图像目标的表示与识别

曹健 著

REPRESENTATION AND RECOGNITION OF THE IMAGE TARGET



机械工业出版社
CHINA MACHINE PRESS

VISIT…

LANZAROTE Caliente.com

图像目标的表示与识别

曹健　著



机械工业出版社
CHINA MACHINE PRESS

本书较为全面地介绍了图像目标识别的相关概念、原理和技术方法。主要内容包括图像目标的特征提取、图像目标的表示与描述、图像目标匹配和图像目标分类等。本书紧跟上述内容的国内外发展现状和最新成果，阐述了作者对图像目标识别的理解和认识，尤其针对局部特征在图像目标识别中的应用，进行了深入的探讨、分析和实例验证。

本书可以作为从事图像理解、模式识别、机器视觉等相关专业研究人员的参考书，对于计算机科学与技术、信息与通信工程、电子科学与技术等专业的研究生和高年级本科生也有一定的参考价值。

图书在版编目（CIP）数据

图像目标的表示与识别/曹健著. —北京：机械工业出版社，2012.7
ISBN 978-7-111-38182-2

Ⅰ.①图… Ⅱ.①曹… Ⅲ.①图像处理 Ⅳ.①TN911.73

中国版本图书馆 CIP 数据核字（2012）第 081984 号

机械工业出版社（北京市百万庄大街 22 号　邮政编码 100037）
策划编辑：吕　潇　责任编辑：吕　潇
版式设计：霍永明　责任校对：薛　娜
封面设计：赵颖喆　责任印制：杨　曦
北京中兴印刷有限公司印刷
2012 年 6 月第 1 版第 1 次印刷
169mm × 239mm · 13.25 印张 · 235 千字
0001— 2500 册
标准书号：ISBN 978-7-111-38182-2
定价：39.80 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务
社服务中心：（010）88361066
销售一部：（010）68326294
销售二部：（010）88379649
读者购书热线：（010）88379203

网络服务
门户网：http：//www.cmpbook.com
教材网：http：//www.cmpedu.com
封面无防伪标均为盗版

前　言

图像目标的表示与识别作为图像处理与模式识别领域的一个重要的研究方向，在安全监控、军事侦察、产品检验、人机交互、医疗诊断等方面得到了越来越广泛的应用。但目前尚未形成一个成熟统一的技术方案，往往需要针对特定的任务，甚至针对特定的图像，选用一种或几种不同的方法。而相关领域的数学算法和具体技术林林总总各不相同，甚至从思路上就已经大相径庭，这更需要我们下工夫进行梳理和提炼。

本书围绕着图像目标的表示与识别这一主题，全面系统地介绍了相关的概念、原理和技术方法。针对可见光图像和刚性目标，学习并借鉴了图像工程、模式识别、机器视觉和人工智能学科中一些先进技术，探讨了复杂背景下的目标识别以及局部遮挡物体的识别中的关键问题，为增强现有图像识别系统的自动化程度和信息处理能力提供理论支持和技术帮助。

本书分为7章，内容安排如下：

第1章概述了图像目标识别的基础理论和研究思路，介绍了图像目标识别常用的图像库，指出了图像目标识别的主要难点和发展趋势；第2章讨论了图像分割和目标分割的关系，介绍了提取目标整体特征的相关技术；第3章介绍了目标匹配和目标分类的基本理论，详细论述了常用的图像目标分类器的设计和训练方法；第4章回顾了局部特征的研究现状，给出了几种典型区域检测算子和特征描述子的具体算法和改进方法；第5章针对局部特征匹配在目标图像拼接和图像检索中应用的不足，提出了基于多分辨率技术的航拍图像拼接方法，以及基于原型匹配的图像检索方法；第6章阐述了视觉单词的理论依据以及视觉单词库特征库构造方法，结合信息论的相关技术进行特征选择，提出了一种基于局部特征的目标分类方法；第7章结合主分量法和Hausdorff距离，提出了一种在视点变化下目标匹配识别方法和基于角点标记图的BP网络分类方法。

本书的研究成果首先要感谢北京理工大学计算机学院的刘玉树教授等多位老师给予作者的长期指导和教诲，还要感谢众多师兄弟和一些硕士研究生在作者攻读博士学位期间给予的启发与激励，更要感谢北京工商大学计算机与信息工程学院的领导和同事们不遗余力的关怀和帮助，尤其感谢国家自然科学基金

和北京市自然科学基金项目（编号：4123095，4112016）对本书相关课题研究的支持。本书的出版也得到了北京工商大学青年教师科研启动基金资助项目（编号：QNJJ2011-38）以及实验室课题组项目（科技创新平台，编号：19005118053）经费的支持，在此一并致谢。

由于图像处理与模式识别领域的相关技术仍处于不断发展和完善阶段，加之作者水平有限，书中难免存在一些不足之处，敬请读者批评指正。

曹　健

2012 年 5 月

目　　录

前言
第 1 章　绪论 ………………………………………………………… 1
1.1　引言 ………………………………………………………… 1
1.2　图像目标识别的意义 ………………………………………… 2
1.3　图像目标识别的框架与思路 …………………………………… 5
1.3.1　图像目标识别问题的分类 ………………………………… 5
1.3.2　图像目标识别的基本框架 ………………………………… 7
1.3.3　图像目标识别的两种思路 ………………………………… 8
1.4　图像目标识别的数据集 ……………………………………… 10
1.5　图像目标识别的开发环境 …………………………………… 15
1.6　主要难点与发展趋势 ………………………………………… 18
1.7　研究内容与结构安排 ………………………………………… 21
1.7.1　本书的研究内容 ………………………………………… 21
1.7.2　本书的结构安排 ………………………………………… 22
第 2 章　图像目标的整体特征提取 ……………………………… 25
2.1　引言 ………………………………………………………… 25
2.2　图像目标分割 ………………………………………………… 29
2.2.1　图像目标分割概述 ……………………………………… 29
2.2.2　图像目标分割现状 ……………………………………… 30
2.2.3　图像目标分割技术 ……………………………………… 33
2.3　目标的表示与描述 …………………………………………… 42
2.3.1　光谱特征 ………………………………………………… 42
2.3.2　纹理特征 ………………………………………………… 44
2.3.3　形状特征 ………………………………………………… 46
2.4　特征空间的优化 ……………………………………………… 48
2.4.1　特征选择 ………………………………………………… 48
2.4.2　特征变换 ………………………………………………… 50
2.5　本章小结 ……………………………………………………… 52

第 3 章　基于整体特征的目标识别 …… 55
3.1　引言 …… 55
3.2　模式识别方法概述 …… 56
3.3　目标匹配的研究现状 …… 58
3.3.1　两种目标匹配方式 …… 58
3.3.2　匹配的相似度度量 …… 59
3.4　目标分类的研究现状 …… 61
3.4.1　分类器设计技术 …… 62
3.4.2　性能评估方法 …… 64
3.5　典型的图像目标分类器 …… 66
3.5.1　基于聚类分析的分类器 …… 66
3.5.2　基于朴素贝叶斯的分类器 …… 69
3.5.3　基于 BP 神经网络的分类器 …… 71
3.5.4　基于支持向量机的分类器 …… 73
3.6　本章小结 …… 76
第 4 章　图像目标的局部特征提取 …… 77
4.1　引言 …… 77
4.2　特征区域的稀疏选取算法 …… 78
4.2.1　特征区域检测的研究现状 …… 78
4.2.2　高斯差分检测算子 …… 80
4.2.3　边缘点检测算子 …… 83
4.3　局部特征的定量描述 …… 85
4.3.1　特征区域描述的研究现状 …… 85
4.3.2　基于梯度分布的描述子 …… 87
4.3.3　线矩特征描述子 …… 89
4.4　角点的检测算法 …… 90
4.4.1　直线投影检测算法 …… 91
4.4.2　SUSAN 算法的自适应阈值改进 …… 92
4.5　实验结果与分析 …… 94
4.6　本章小结 …… 97
第 5 章　基于局部特征的目标匹配 …… 99
5.1　引言 …… 99
5.2　结合 NNDR 与霍夫变换的匹配方法 …… 100
5.2.1　基于 NNDR 的匹配策略 …… 100
5.2.2　邻近特征点的搜索算法 …… 101

5.2.3　基于霍夫变换的目标检测 …… 103
5.3　基于局部特征和多分辨率技术的图像拼接 …… 105
5.3.1　图像拼接技术的研究现状 …… 105
5.3.2　多分辨率下的图像配准 …… 107
5.3.3　渐入渐出的图像融合算法 …… 112
5.4　基于局部特征和原型匹配的图像检索 …… 114
5.4.1　CBIR 的研究现状和发展趋势 …… 114
5.4.2　基于模板匹配的检索方法 …… 117
5.4.3　基于原型匹配的反馈技术 …… 118
5.5　实验结果与分析 …… 119
5.6　本章小结 …… 124
第 6 章　基于局部特征的目标分类 …… 127
6.1　引言 …… 127
6.2　目标的向量空间模型表示 …… 129
6.3　构造视觉单词库 …… 130
6.3.1　视觉单词的生成方法 …… 131
6.3.2　基于 RNN 的层次聚类算法 …… 132
6.4　基于信息论的特征选择方法 …… 134
6.4.1　信息论的相关概念 …… 135
6.4.2　基于信息增益法的特征选择 …… 136
6.4.3　基于 CHI 统计量的特征选择 …… 137
6.4.4　基于互信息法的特征选择 …… 138
6.5　视觉单词的权重计算 …… 139
6.6　实验结果与分析 …… 141
6.7　本章小结 …… 146
第 7 章　基于角点特征与视面模型的目标识别 …… 147
7.1　引言 …… 147
7.2　三维物体的视面模型表示 …… 150
7.3　基于角点特征的目标匹配 …… 152
7.3.1　利用基准角点进行目标匹配 …… 152
7.3.2　基于主分量与 Hausdorff 距离的匹配算法 …… 154
7.4　基于角点标记图的目标分类 …… 157
7.4.1　角点特征的优化技术 …… 157
7.4.2　角点标记图的生成方法 …… 159
7.5　实验结果与分析 …… 160

7.6　本章小结 …… 164

附录 A　图像处理的一些相关理论 …… 167

A.1　数字图像的基本概念 …… 167

A.2　数字图像的信息内容 …… 168

A.3　图像处理的技术门类 …… 169

附录 B　模式组合的一些基本概念 …… 173

B.1　图 …… 173

B.2　树 …… 173

B.3　符号串 …… 174

附录 C　概率统计的一些预备知识 …… 177

C.1　概率 …… 177

C.2　最大似然估计 …… 177

C.3　条件概率 …… 177

C.4　贝叶斯公式 …… 178

C.5　随机变量 …… 179

C.6　二项式分布 …… 179

C.7　联合概率分布和条件概率分布 …… 179

C.8　贝叶斯决策理论 …… 180

C.9　期望和方差 …… 181

附录 D　信息检索的一些基础模型 …… 183

D.1　布尔模型 …… 183

D.2　向量空间模型 …… 183

D.3　概率模型 …… 184

D.4　语言模型 …… 185

附录 E　名词术语解释 …… 187

参考文献 …… 192

第 1 章　绪　　论

我们只能向前看到很短的距离，但是我们能够看到仍然有很多事情要做。

——阿兰·麦席森·图灵（1912—1954）

1.1　引言

视觉是人类获取信息、感知世界，进而改造世界的一个重要途径。有资料显示，人类接受到的外界信息中约有 60% 以上来自于视觉，而听觉、味觉、触觉、嗅觉总共占不到 40% 。但是从技术发展来看，图像信息的处理远远滞后于语音信息，随着计算能力的不断提高，如何使计算机具有和生物类似的视觉感知功能成为目前计算机领域中的一个研究热点。

图像目标的表示与识别，又称关于视觉图像的模式识别，旨在利用图像处理与模式识别等领域的理论和方法，确定图像中是否存在感兴趣的目标，如果存在则为目标赋予合理的解释，必要时还要确定其位置[1] 。虽然国内外科研工作者就如何在复杂环境下检测、辨识和准确跟踪目标进行了理论分析和实践探索，但目前尚未形成一个成熟统一的技术方案，往往需要针对特定的任务，甚至针对特定的图像，选用一种或几种不同的方法。而相关领域的数学算法和具

体技术林林总总各不相同，甚至从思路上已经大相径庭，这更需要我们下工夫进行梳理和提炼。

在这里，识别（Recognition）、分类（Classification）、检测（Detection）、定位（Location）和鉴别（Identification）几个概念需要简要说明一下。从上面的定义可以看出，识别的内涵最为宽泛，分类、检测、定位、鉴别都能看做是识别的子任务之一；分类的定义比较清晰，即对图像目标按照类别标签进行划分；检测和定位的目的是相似的，一般是确定图像中某个目标的具体位置；鉴别往往指同类目标间的区分，如对人物张三和人物李四进行辨认。虽然这几个概念在不同的文献中稍有差异，本书中对它们的解释也并不唯一，然而把握好它们在具体问题中的界定，还是有助于加深对图像识别领域中实际问题的理解。

1.2 图像目标识别的意义

近年来，许多重要的国际期刊（IEEE Transactions on Pattern Analysis and Machine Intelligence、IEEE Transactions on Image Processing、IEEE Transactions on Medical Imaging、IEEE Transactions on Vehicular Technology、International Journal of Computer Vision、Computer Vision and Image Understanding、Image and Vision Computing、Pattern Recognition、Pattern Recognition Letters、Machine Vision and Application 等）以及重要的国内期刊（计算机学报、软件学报、自动化学报、机器人、模式识别与人工智能、计算机研究与发展等）都发表了大量关于图像模式识别方面的论文。在国外召开的顶级国际会议，如 IEEE 国际计算机视觉与模式识别（Computer Vision and Pattern Recognition，CVPR）会议、欧洲计算机视觉会议（European Conference on Computer Vision，ECCV）、国际信息处理会议（International Conference on Information Processing，ICIP）等，也收录了许多知名学者在相关领域的学术成果。这几年，国内学术界积极开展了一系列的学术交流活动，比如 2005 年在北京举办的国际计算机可视化会议（International Conference on Computer Vision，ICCV）、2006 年在香港特别行政区举办的第 18 届模式识别会议（International Conference on Pattern Recognition，ICPR）、2008 年全国模式识别学术会议、2009 年在西安举办的第 9 届亚洲计算机可视化会议（Asian Conference on Computer Vision，ACCV）等。

图像目标的表示与识别之所以备受关注，是由于它能够广泛应用于国防和民用的许多领域，其中包括安全监控、军事侦察、产品检验、人机交互和医学

应用等多个方面。

1. 安全监控

图像目标识别在安全领域的应用范围非常广泛，大城市很多地方，如民宅、停车场、银行等，都装有闭路电视监控系统（Close Circuit TV），以便能够对可疑的物品和人员进行有效的监控。而随着各种新的 DNA 分析、分型技术方法的建立，借助多模态的生物特征辨识系统，法医 DNA 分析技术可将从犯罪现场提取的 DNA 轮廓（手掌的纹理、指纹、脸的几何形状）与疑犯的 DNA 信息进行更准确、快速、自动的匹配。2004 年，根据市场研究公司——国际生物测定组织的分析，在人们首选的在线银行认证方法中，选择生物特征辨识的占了 50%，是智能卡、密码、身份证号码等方式的总和。

在交通系统中除了视频摄像外，还需要大量的识别监视跟踪系统。例如，目前的车牌识别技术已经非常成熟，这对道路上异常车辆的监控和交通事故的事后处理都具有非常重要的意义。西门子公司的交通监控性能非常优越，不仅能探测隧道中慢行或停止的汽车，还可探测处于 U 形转弯处的违规汽车，以及自动检测可疑的行李。智能车辆的最终目的是实现车辆的自动驾驶，目前主要是利用车上安装的摄像机、雷达等传感器设备进行道路检测并识别前方的障碍物（如车辆、行人），以保证车辆的安全行驶。

2. 军事侦察

相对而言，军事领域的识别与监测要求就非常苛刻了，主要是因为战场环境要比一般的民用环境更为复杂。例如，检测有遮挡和伪装的机动目标就十分困难，由于假设的局限性，在民用上已经比较成熟的算法在军事上往往效果很不理想。美国洛克希德·马丁公司开发的数字式侦察图像处理系统已安装到尼米兹级航空母舰上，成为美国海军联合部队图像处理系统（JSIPS-N）的战术组成部分，它能接受和处理来自多个传感器平台（U-2、“全球鹰”无人机、F/A-18 共享侦查吊舱等）的图像，极大增强了美国海军识别和打击关键目标的能力。2006 年 6 月以色列 IAI 公司在巴黎展示了其一元化的战争指挥室，其中实时图像情报中心（EL/S-8894RT-RiCENT）具有对战场全天候一体化的监视和侦察能力。

3. 产品检验

由于工业环境的结构、照明等因素可以得到严格的控制，图像目标识别在工业生产和装配中得到了成功的应用。一个具有简单视觉感知功能的自动化生产线包含一个摄像机和相关的信息处理系统，通过摄像机对零件进行识别和定位，为机器人提供是否操作或进行何种操作的信息，并引导机器人手臂实时准

确地夹取零件；此外，图像识别技术已经应用在集成电路设计、图形设计和电视电影制作中；通过多源图像融合，可以进行产品外形检验、表面缺陷检验，加强对产品质量的严格把关。

对多个摄像机的图像同步识别处理，利用某一时刻关于某个目标的不同角度的图像可以恢复场景的三维信息，并依据三维信息做出决断，实现即时规划、自主导航、与周围环境实时交互作用等。这是生产控制的进一步发展，让机器人不仅仅停留在简单的自动化生产线上，而且能够代替人类进入危险的环境进行生产活动，例如，在核辐射区或火灾现场抢修设备，远程控制的无人开采矿藏、星际探测设备的自主导航等。

4. 人机交互

对包含文字和符号的图像进行识别可以让人与计算机的交互更加便捷。目前这方面的技术大量应用于信函分拣、稿件输入、支票查对、期刊阅读和自动排版中，而超市的条码阅读器更是对销售管理的一场革命。现在美国和日本的客户已经能够通过把他们的手机指向汉堡包的包装纸，获得其营养信息并显示在屏幕上，也可以通过这种方式获得商品报价。例如，在日本东京的一座建筑物上粘贴的超高速识读条码（Quick Response Code，QRcode）就含有很多信息，通过带有摄像头的可正确编译的手机就能方便地读取。

面部表情传达了一种非口头性的暗示，对其进行自动识别是人机接口的重要元素，也被用于行为科学和临床实践中。比如，具有微笑探测和眨眼探测的两个功能独立的数字照相机可以在恰当的时机（用户微笑的时候或眨眼之后）捕捉到主体，并提示用户，进行抓拍。手势识别也称手语识别，是机器视觉领域中比较前沿的研究领域。当用户做出一个手势，摄像机（一般为双目或三目）将图像传送到计算机，然后由特定软件结合视差来提取手臂、手指等三维特征，完成这些特征的进一步识别，最后对这个手势做出响应[2]。

5. 医学应用

如今，计算机图像分析逐步融入到了医疗诊断的过程中，这就促生了计算机辅助诊断（Computer-Assisted Diagnosis，CAD）技术。利用该技术，可进行核磁共振成像（主要用于医疗成像来可视化人体结构和功能，提供任何平面内身体的细节图像），癌细胞、白细胞、染色体检查，修复手术控制设计等。

通过一组切片图像进行人体器官的三维重构，可以为医疗诊断和病理分析提供重要和直观的帮助。同样，可以根据图像序列中的信息对普通目标进行三维重构，无论观察点在何处，都能利用其三维信息进行识别，这也为解决视点

变化下的目标识别提供了一个思路。

除了以上几个方面，图像目标识别在生产生活中还有很多应用。对目标描述信息的分析处理，可以用在天气预报、森林火灾及地质灾害监测、空气污染预报等领域。人脸检测（Facial Detection）技术可以将画面及时地锁定在讲话人身上，这样就很大程度地降低了远程电视会议的图像传输比率[3]。在虚拟现实、计算机动画、视频评注等应用领域，目标识别技术同样也发挥着不可替代的作用。所以，开展图像目标识别研究意义重大，其研究成果具有非常广阔的应用前景。

1.3　图像目标识别的框架与思路

1.3.1　图像目标识别问题的分类

针对图像领域中的各种具体问题，目标识别所采用的研究方法和技术方案都有所区别，甚至迥然不同。所以需要将目标识别问题按照一定的标准进行分类，对具体问题进行具体分析。

1. 按照获取图像的传感器的种类

按照获取图像的传感器的种类，可以将图像目标识别分为可见光图像目标识别，红外图像目标识别和合成孔径雷达（Synthetic Aperture Radar，SAR）图像目标识别。这三种传感器的成像原理不同，对拍摄时间、天气情况、地理环境、光照的要求也不一样。通常条件下，可见光图像比较清晰、直观、费用低，有利于实时传输，但可见光传感器只敏感于目标场景的可见光反射，容易受到各种场地因素的干扰；红外图像[4]适合夜间使用，具有特殊的识别伪装的能力，但图像清晰度低，且大气红外辐射和吸收作用对图像质量影响很大；合成孔径雷达图像[5,6]易于判读线性地物、表面光滑的面状地物、森林、草地、水体等，具有很强的穿透力，但雷达视向对目标的表达色调和形状影响很大。目前，国外先进的无人侦察平台都采用多种传感器成像技术，并通过图像融合得到了信息更为丰富的图像。

2. 按照图像背景的复杂程度

按照图像背景的复杂程度，可以分为简单背景下的目标识别和复杂背景下的目标识别。简单背景下的目标识别，如文字识别、符号识别和人脸识别等，目标和背景的对比度非常大，一般的图像处理和分割算法就能准确完整地提取出目标。此类研究侧重于如何辨识出更加细微的区别，或者对目标的不同姿态进行识别。而在复杂背景下进行图像目标识别受到噪声的影响非常大，目标的

检测效果往往差强人意，要想提取出完整的目标更是困难，一般需要在先验知识的指导下进行目标的检测和图像的分割。

3. 按照相关图像的性质

按照相关图像的性质，可以分为静态图像识别和动态图像识别。静态图像，也称静止图像，指的是关于目标的单幅图像，我们一般的图像检索和图像分类大都是针对这类图像的。而动态图像为我们提供了比静态图像更为丰富的信息，通过对多帧动态图像（图像序列）的分析，可以检测出目标的运动信息，识别与跟踪运动目标和估计三维运动及结构参数。动态图像识别面临的首要挑战是，如何从图像序列中实现有效的图像分割和图像对应。图像分割在静态图像识别领域也尚未得到有效解决，图像对应问题则是与模式识别和人工智能紧密相连的难题。

4. 按照图像中目标的数目

按照图像中目标的数目，可以分为单目标识别和多目标识别。单目标的图像，顾名思义，就是只有一个感兴趣的目标，其余属于背景，这就相当于提供了一个重要的前提。在这个前提下，我们更多关注的是如何利用各种图像处理技术抑制背景，完整准确地检测和提取出这一个目标。而多目标识别要比单目标识别困难得多，因为多个目标同时出现在一幅图像中，不光有复杂背景的干扰，还必须考虑到目标之间会相互遮挡（Occlusion）、合并（Merge）、分离（Split）等种种情况。这更需要通过知识来指导信息的选择和整合，并进行反复的假设验证（Hypothesis Verification）和复杂的反馈处理。

5. 按照图像中目标的类型

按照图像中目标的类型，可以分为刚性目标识别和非刚性目标识别。刚性（Rigid）目标一般指具有刚性结构、不易变形的物体，如飞机、车辆、建筑物等人造物体，它们的共同特点是结构比较规范，适合用几何模型进行描述，一般采用基于形状特征的方法进行识别。而非刚性（Non-rigid）是指外形能够变化的物体，如细胞、动物、人体等。对这类目标可以采用光谱特征、纹理特征以及变形模板（Deformable Template）技术等进行识别。

6. 按照对图像语义的理解程度

按照对图像语义的理解程度，可以分为图像分类、目标检测以及目标识别。图像分类只是根据低层图像特征和相似度度量，将内容类似的图像归为一类，并不需要对图像中的对象进行分割和定位，如基于内容的图像检索；目标检测不仅要确定图像中是否存在感兴趣的目标，还要在必要时确定其位置，以便于

进一步提取目标进行处理，如车牌提取、人脸检测等；具体目标的识别需要对图像信息进行深入分析，例如在视觉跟踪中，不仅要检测出感兴趣的目标，还要与周围其他目标进行区分，避免产生混淆。

1.3.2 图像目标识别的基本框架

一个典型的图像目标识别系统如图 1-1 所示，主要由图像增强与变换（图像预处理）、图像分割、图像描述、分类决策四部分构成。这四个部分关系非常密切，在看做一个有机整体的同时，也可以看成三个层次的计算处理——低层、中层和高层处理。

低层是对图像数据进行预处理，如对有噪声的图像要进行滤波去噪，对信息微弱的图像要进行对比度增强，对失真图像要进行几何校正等，以达到改善图像质量、突出兴趣区域的目的。其鲜明的特点是输入和输出的都是图像。

中层处理涉及分割（把图像分为不同区域或目标物），将给定图像或已分割的图像区域用更为简单明确的数值、符号或图来表征（特征描述），以使其更适合计算机处理及对不同目标的分类（识别）[7]。中层处理输入为图像，但输出的是从这些图像中提取的特征。

高层处理一般是基于知识进行推理和证实的，涉及图像或图像区域的理解，以及执行与视觉相关的识别函数[8]。也可以简单地认为是对图像或图像区域进行分类和估计。其输入是向量、串或树等形式的特征描述，输出则是图像或物体的类别。

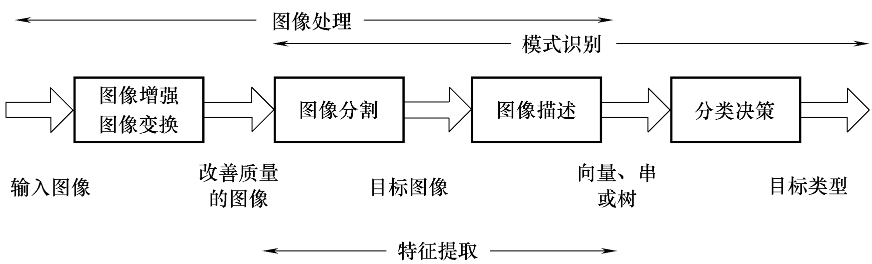


图 1-1 图像目标识别系统的基本框架图

图像目标识别技术是延伸和扩展人的视觉功能的方式和方法，其实信息技术整体都可以认为是扩展人的信息器官功能的技术。人的信息器官主要包括感觉器官、传导器官、思维器官和效应器官四大类型，其功能主要是信息获取、信息传输、信息处理和信息应用，因此感测技术、通信技术、智能技术与控制

技术被认为是信息技术的四基元，其他信息技术通常被看做是这四种基本技术的高阶逻辑综合或分解衍生[9]。

如表 1-1 所示，我们把图像识别看作图像处理和模式识别的交叉，而这两门学科分别属于信息处理和计算智能两个大的学科门类，甚至还涉及信息传输的一些内容，从这个意义上，也看出设计和执行算法来模仿人类对物体的视觉识别能力是一项有趣而富有挑战性的任务，因此，这门学科不断吸引了许多来自不同领域的科研人员钻研和探讨，也不断涌现新的理论和方法。

表 1-1　图像识别在信息学科中的位置

信息器官（人）	器官的作用	相 应 技 术	学 科 门 类	研究方向（举例）
感觉器官	信息获取	感测技术	信息处理	图像处理 信号分析
传导器官	信息传输	通信技术	信息传输	信息编码 信息安全
思维器官	信息加工	智能技术	计算智能	人工智能 模式识别
效应器官	信息应用	控制技术	自动控制	集中控制 分散控制 现场控制

1.3.3　图像目标识别的两种思路

人类认知过程可以用图 1-2 描述[10]。不同视觉基本特征，如方位、方向、空间频率、眼优势、空间拓扑和颜色等在不同层次视觉皮层具有一定的空间组织形式，多种基本特征功能柱共存于一片皮层空间，实现多种特征表达的最优化；特异性反应细胞在高级与初级视觉皮层上进行自下而上的前馈和自上而下的反馈，完成视觉表征自下而上地逐级抽象，以及在整合后自上而下地反馈、对初级水平的调控；大脑自动建立基于皮层自组织的计算视觉模型[11,12]。

对于图像目标识别问题的研究，也是遵循着人的认知形式，总体上讲有两种思路，一种是自下而上的加工（Bottom-up Process），另一种是自上而下的加工（Top-down Process）[13]。这两类处理方法有着各自的优点和缺点，将它们结合起来各取所长，就有可能实现更为理想的识别。

1. 自下而上的加工

也被称为数据驱动（Data-driven）的加工，其核心观点是系统工作是单向

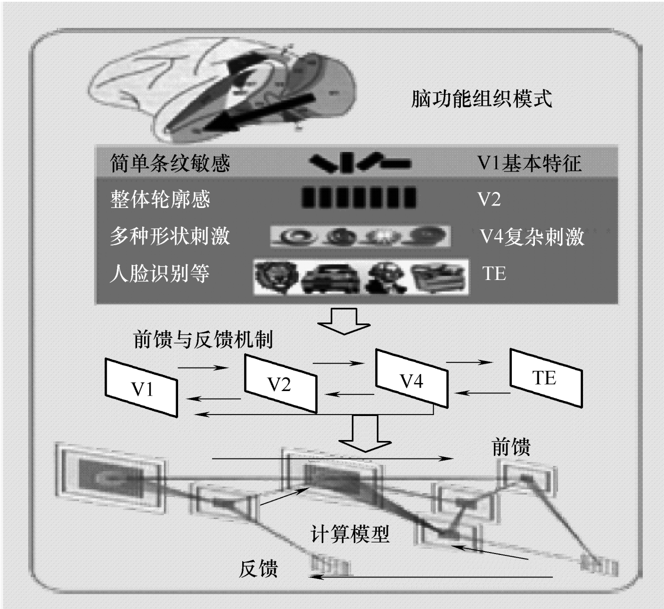


图 1-2　人类的认知过程

的，从信息输入开始，一直到形成最终的解释。无论在特定的时刻发生什么都不受后面加工过程的影响，这种加工系统无法回到先前的阶段去调整[13]。反映在图 1-1 上，就是严格按照先后顺序，从低层开始处理图像上的数据，到中层处理将这些数据转化为抽象表征，到高层进行识别，其间各个模块互不相关。

自下而上的视觉计算理论中，马尔模型[14]显然最具代表性，它在技术性和数学形式方面堪称精彩。马尔认为，视知觉是通过构建三种不同心理表征或素描进行的。首先是原始素描，它以二维图像的方式描述相对明暗的区域和已经固定位置的几何结构，使得观察者能够分辨不同区域的边界，但无法“得知”这些视觉信息的“涵义”；然后建立一个更为复杂的表征，即 2.5D 素描，观察者利用阴影、纹理和边界等线索，获得关于该素描表面的信息，以及此刻它们在景深上与观察者的相对位置关系；马尔认为原始素描和 2.5D 素描所依据的都是数据驱动，只有在观察者最后构建视觉场景的三维素描时，有关现实世界或特定期望的信息（知识）才会被纳入进来。

这类方法的优点是便于工程实现，对单目标识别及复杂图像分析系统均适用，具有较强的代换性，现有许多系统在解决图像识别方面的问题时都遵循这一思路；缺点是在分割、标记、特征提取等处理过程中缺乏知识指导，盲目性大，因此在很大程度上局限了该方法的应用范围。

2. 自上而下的加工

又称之为理论驱动（Theory-driven）或概念驱动（Conceptually Driven）的加工，知识理论或概念假设引导系统在识别过程中的信息选择和整合。其基本思想是利用先验知识为待识别问题建立模型，然后利用图像数据对模型的正确性进行验证，此类方法有坚实的数学理论基础，有许多数学工具可以使用，因此一直是理论界研究模式识别问题的主流方法。

先验知识可以通过专家总结、人为定义来得到，有了待识别目标的模型特点，就可以在图像中进行有针对性的检测和识别了。这种方法一般用于线状目标和形状规则的刚性目标的识别[1,15-17]，如机场、道路、门窗等。但是人的很多先验知识很难用数学形式来表达，所以近年来随着机器学习技术被大量应用于机器视觉领域，用统计学习的方法获取隐含的知识模型已经成为了研究热点[18-22]。

自上而下的加工方法，其优点在于底层处理是在知识指导下的粗匹配过程，可避免抽取过多不必要的特征集，提高算法的效率，其精匹配过程也因而变得简单和有针对性。它的缺点是代换性和兼容性差，识别目标改变，知识和假设要随之而变。

1.4 图像目标识别的数据集

图像目标识别系统的实验比较和性能评估往往是在一些标准图像库上进行的，描述规整、功能强大的图像数据集对于图像目标识别过程以及评价体系的建立非常重要。图像数据与一般事务数据不同，它的数据量大，具有多维性和多样性。

根据模式识别的理论知识，如果一个训练数据集代表了对象集的总体分布，那么识别系统对新的对象操作的性能就和对训练数据集一样。然而，获取足够大的数据集经常是一件费力的事。为了使数据集成为具有代表性的，它必须包括可能遇到的各种类型对象的例子，包括一些不常见的对象[23]。

一些传统的图像数据集，如 Caltech、PASCAL 等，为图像目标识别学习评价提供了基线标准，这些图像库在理想环境下评价算法的性能相对较高，但缺乏背景变化复杂的真实图像，因此无法评价图像目标识别方法的健壮性和自适应性。新兴的图像数据集，如 LabelMe、LotusHill 等，基于特定的图像识别任务，从简单的图像分类到海量图像检索和网络图像注释，已经融入了基本的视

觉数据和相应的先验知识。

1. Caltech 图像库

Caltech 图像库含有 Caltech-101[㊀] 和 Caltech-256[㊁] 这两个数据集。Caltech101 是由加州理工学院的 Li 等创建的图像集，有 101 类目标，每类目标有 40 ~ 800 幅图像，图像大小约 300×200 像素，并对每幅图像都进行了注释，每个注释包括两种信息：一是目标位置的边界盒，二是人工描绘的目标轮廓。

Caltech-101 图像库的优点在于图像大小和目标相对位置大体相同，不需要花时间去裁剪图像就能进行实验；图像的杂乱或遮挡部分很少，识别算法可以依赖于目标的少数特征；对目标轮廓的细节进行了注释。其缺点是图像目标种类较少，真实世界的目标粗略分类也达到万以上的数量级；图像大都过于简单，而图像目标通常在相对位置和方向上有更多的变化，目标和背景之间也会存在遮挡。

2007 年，加州理工学院在 Caltech-101 基础上又创建了 Caltech 256 图像库。该图像库包含了 30607 幅图像，共分为 256 类目标，类别相当于原来的两倍多。各类图像的最小数目增至 80 幅，还加入了更复杂的图像背景和更多的目标姿态变化。

2. Corel 图像库

Corel 是基于不同场景的图像库，共 6 个大类，平均每类有近 14000 幅图像，广泛应用于基于内容的图像检索领域。它是注释后的图像集，但注释是不同的研究人员完成的，注释质量差别很大，由于注释的多变性，目前无法直接进行图像目标识别的性能估计。此外，不同的研究课题通常采用不同场景的图像库子集进行测试，所以很难对基于该图像库的目标识别结果进行直观的比较。

美国宾夕法尼亚州州立大学的王教授从 Corel 标准测试图像库中挑选出来一个子集，称为 WANG 图像库[㊂]，被广泛应用于对识别效果验证。该图像库中包含非洲原始居民、海滩、建筑物、公交汽车、恐龙、大象、花卉、马、雪山、食品 10 类共计 1000 幅彩色图像，皆存储为 JPEG 格式，大小为 256×384 像素或 384×256 像素。如果查询图像来自于 10 类中的一类，查询者可以从此类中找出其他图像，有利于评价检索结果，当然，该图像库也可以用来评价图像分类的效果。

㊀ http：//www. vision. caltech. edu/Image_ Datasets/Caltech101。

㊁ http：//www. vision. caltech. edu/Image_ Datasets/Caltech256。

㊂ http：//bergman. stanford. edu/ ~ zwang/project/imsearch/WBIIS. html。

3. COIL 图像库

COIL 图像库是美国哥伦比亚大学计算机科学学院创建的图像数据集。COIL-20 里含有 20 种不同 3D 目标的 72 种视角（以 5°为间隔）灰度图像。每幅图像包括一个单一目标，且在不同光照条件下，这些目标在均匀的黑色背景前，共有 256 个灰度级。总计有 1440 幅大小为 128×128 像素的参考图像，称为“处理过的数据”（Processed Data），还有 360 幅大小为 448×416 像素的测试图像，称为“未处理的数据”（Unprocessed Data），这两类图像的光照条件和大小都不同，可以满足训练和测试相互独立的要求。

为了增大图像识别的难度，在 COIL-20 的基础上又创建了新图像库和新背景图像库，每个测试集都有目标的转换参数信息，可以将不同转换的影响分开。Keysers 在 2006 年创建了两种不同背景的图像库：COIL-RWIH-1 和COIL-RWIH-2。前者包括在均匀背景中的目标，后者包括现实世界中不均匀背景中的目标，它们也都分为训练图像和测试图像，并且分辨率不同。COIL-100 图像库包含 100 个目标的 7200 幅彩色图像（平均每个目标 72 幅），目标有多变、复杂的几何和影像特性。

4. PASCAL 图像库

PASCAL（Pattern Analysis，Statistical Modeling and Computational Learning）图像库是 2005 年由欧洲的苏黎世大学、爱丁堡大学及牛津大学组织倡导的，由相应的专项基金支持，旨在构建含有海量数据的公用图像库，在现实场景中识别多个目标类别信息，为全世界的图像识别研究人员提供一个基准，进行相应的算法分析和方法比较。PASCAL 视觉目标识别竞赛（从 2005 年开始，每年一次）也采用该图像库，这个图像库包含标注信息，是目前识别难度最大的数据集之一，而且每年都进行类别和数量的扩充，并做相应的技术统计报告。

PASCAL2005㊀包含 4 类目标（摩托车、自行车、汽车和人）在不同姿势、不同视角下的照片；PASCAL2006㊁包含 10 类目标（自行车、小汽车、摩托车、人、公共汽车、猫、狗、母牛、马、绵羊）共 5304 幅图像，都标注了位置（目标边界框）及类别名称；PASCAL 2007㊂中共包含训练图像 2501 幅，验证图像 2510 幅，测试图像 4952 幅，包括自行车、小汽车、摩托车、公共汽车、船、火车、飞机、人、猫、狗、母牛、马、绵羊、鸟、植物、瓶子、餐桌、沙发、椅

㊀ http：//www. pascal - network. org/challenges/VOC/voc2005/index. htm。
㊁ http：//www. pascal - network. org/challenges/VOC/voc2006/index. htm。
㊂ http：//www. pascal - network. org/challenges/VOC/voc2007/index. htm。

子、显示器20个类别，这些真实场景中的图像中可能同时包含几类目标，目标的大小比例变化很大，检测目标存在遮挡、变形，同类目标之间也有较大的差距，每幅图像有相应的按规范格式书写的标注文件，标明了图像中包含的目标名称、边界盒、视点（前视图、后视图、左视图、右视图、未知视图）及识别难易度；PASCAL 2008㊀的目标类型和PASCAL 2007没有太多变化，同样是20类，只是多了一些分割的标注信息，另外，难度也有所增强。

PASCAL图像库对每幅图像中目标的位置及类别的标注，使得在测试过程中可以分别检验图像分类（目标在测试图像中是否出现）和目标定位（测试图像中每个目标的边界框）的效果。PASCAL图像库的另一个特别之处在于提供了两种测试集：第一种测试集中的数据来源于许多传统的标准图像库，如Caltech图像库（训练集和测试集遵循随机的均匀可变分布，许多算法对该图像库已经达到非常好的实现效果）；第二种测试集可以解决新实例的收集问题，通过不同的图像获取途径，如图像搜索、视频监控、航空拍摄等，在尺度变化、多姿态、复杂背景以及局部遮挡等方面为测试集提供了更加丰富的数据，用以评价算法的泛化能力。

5. LabelMe图像库

LabelMe[24]是MIT计算机系人工智能实验室创建的一个允许在线标记和图像资源共享的通用注释工具。该工具提供多边形绘图、图像查询和浏览图像库等许多功能，图像库和所有的注释都可以免费使用，并且支持几乎所有的浏览器，甚至可以支持Javascript标准图形接口。注释的结果保存在XML文件格式中，这样的注释可植入并容易扩展。

LabelMe是一个不断拓展的标记图像库，有11845幅静态图片，18524组图像序列（每组序列至少存在一个标记目标）。图像库中包含111490个多边形组成的目标区域（2006年年底统计），其中44059个是用在线工具标注的，67431个是离线标注的。其一个重要优势在于包含WordNet，可以在WordNet树的不同级别查询目标。

LabelMe与其他图像数据集的主要区别是：LabelMe中的目标是一类而并非个体信息，识别一个目标的类别信息，不但需要同类的不同个体的多张图像，而且需要不同的观察条件；在真实场景中标记目标，使得目标检测具有很强的背景干扰，适合训练基于复杂背景的图像目标识别系统；高质量、在线标注，

㊀ http://www.pascal-network.org/challenges/VOC/voc2008/index.htm。

不仅保证了资源同享而且更多细节信息（如边界框、多边形或分割掩膜）对目标识别和图像分析非常有帮助；许多不同的目标类和大量不同场景的图像，可以面向更多图像识别的应用场合，通过改变目标种类、场景样式、距离远近、背景复杂度等，在分析不同的环节和参数对识别效果的影响时，是非常有用的；LabelMe 是个公开的图像库，采用许多非版权图像，大多数是利用手提式数码相机拍摄的，也有许多利用网络摄像头获取的视频，具有开放性和动态性。

6. 莲花山图像库

上述图像库局限于仅仅标记了目标的粗糙边界，并不适合精细的区域分割或语义分解。因此在 LabelMe 图像库的基础上，出现了另一种包含更为详尽的视觉知识的图像库——莲花山图像库（Lotus Hill Research Institute Image Corpus）[25]。该图像库是由中国莲花山计算机视觉和信息科学研究院创建的，由全职标注人员用解释图（Parse Graph）的方式对每个图像或目标进行了标注，并按照 WordNet 的标准表示目标、部件的名字和关系。

莲花山图像库到 2008 年为止有 3927130 个位置点，636748 幅图像（视频），而且数目还在不断增加，其中 13 个子集一般作为算法评估的基准，如一般场景、事件和活动、航拍图像、热门目标、一般目标、人脸和姿态、视频剪辑、文字、自然图像的 2. 1D 分层表示等。

莲花山图像库不单纯是图像数据的存储管理和查询检索，而且是基于通用需求标记信息的标识法和组织法，构建的一种新的大型的、通用的、真实的图像数据集，实现了图像理解中信息组织和信息运用的两大基本任务。该图像库通过适当组合注释工具的功能模块，可以完成对图像的任何标记和注释工作，并利用知识库的引导加速这一过程。

随着目标种类的增加、同一类目标之间视觉差别的增大，目标识别研究对图像数据的数量和种类有着更为严格的要求。而大多数图像库都是人工收集并加以标注的，这耗费了大量的人力和物力。近些年，不少科研人员在尝试让计算机自动完成这项任务。Fergus 等人[26,27]使用视觉信息对从网上获得的大量图像数据进行标注；Berg 等人[28]则专注于建立几种动物类的图像数据库，他们使用搜索工具从网上搜索图像，通过狄雷克勒分配技术发掘一系列潜在主题和对应的图像样例；Schro 等人[29]利用贝叶斯理论和支持向量机技术实现了图像数据库的自动收集；Collins 等人[30]为了获得精确和大规模的图像数据集，设计了一种判别性学习方法，能主动在线学习快速分类对象并实现数据库的自动

构建。随着目标识别系统的发展，相信会出现更多更好的图像库和图像数据收集算法。

从模式识别的角度来说，数据集在系统性能评估中的应用方式主要有三种[31]：重替代法，就是使用相同的数据集，先进行训练再进行测试，这种方法非常简便，但测试结果通常是偏于乐观的；坚持把可用的数据集被分成两个子集，一个用于训练，一个用于测试，这种方法最为常用，但缺点是划分子集减少了训练和测试数据集的大小，而且需要人为决定用于训练集和测试集中的样本数目；留一法，循环地以每一个样本为测试对象，而数据集中的其他样本作为训练样本，该方法使用了所有样本的同时维持了训练数据集和测试数据集之间的独立性，但缺点是有很高的计算复杂度。本书的实验中在对图像库的使用上采取第二种方案，即划分出两个独立的子集作为训练集和测试集，它们包含的样本数量比例一般为 8∶2 或 7∶3。

1.5　图像目标识别的开发环境

正如 1.3.2 节所述，图像目标识别系统是采集、表达、分析和识别图像中视觉信息的系统，涉及图像处理、模式识别乃至机器视觉知识的方方面面，因此借助一些开源函数工具，可以更有效、更有针对性地研究图像中各种是知识模型与相应的图像数据处理过程。

开发环境（Software Development Environment）一般是指在基本硬件和宿主软件的基础上，为支持系统软件和应用软件的工程化开发和维护而使用的一组软件，简称 SDE。作为一种软件工具，开发环境能够让科研工作者摆脱自己实现底层代码的繁琐工作，从而提高图像目标识别算法的实现效率，加速相关理论和方法的研究进程；在实际应用中，掌握并能熟练使用一种或几种图像目标识别方面的开发环境，将对开发图像目标识别软件十分有帮助。

1. OpenCV 的优势

OpenCV（Open Source Computer Vision Library）是由 Intel 公司资助的基于 BSD 许可证授权（开源）发行的跨平台计算机视觉库，主要面向商业开发或研究学者，目前由 Willow Garage 公司负责日常维护，它的不断发展对智能信息处理、机器视觉、人工智能、图像识别和认知神经科学方面软件的研发都有非常重要的影响。

OpenCV 的主要特点有：

1）轻量级而且高效，由一系列 C 函数和少量 C + + 类构成，其代码都经过优化，可用于实时处理图像。

2）统一的结构和功能定义。

3）具有良好的可移植性，可以运行在 Linux、Windows 和 Mac OS 操作系统上。

4）可以进行图像/视频载入、保存和采集的常规操作，实现了图像处理和计算机视觉方面的很多通用算法。

5）具有底层和高层的应用开发包和方便灵活的用户接口，同时支持Python、Ruby、MATLAB 等语言编程。

6）提供了面向 Intel IPP 高效多媒体函数库（Integrated Performance Primitives）的接口，可针对 Intel CPU 优化代码，提高程序性能（OpenCV 2.0 版的代码已显著优化，无需 IPP 来提升性能，故 2.0 版不再提供 IPP 接口），如图 1-3 所示。

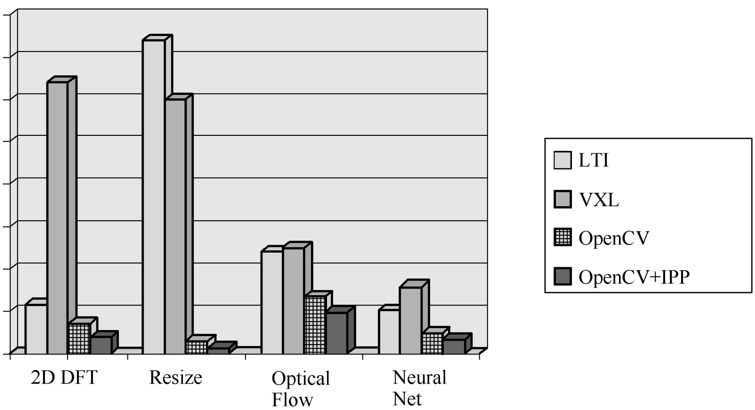


图 1-3　OpenCV 与其他视觉函数库的性能比较

图中所用的视觉函数库版本分别为 OpenCV 1.0 测试版，IPP（Intel Integrated Performance Primitives）5.0，LTI1.9.14 和 VXL（Vision Something Libraries）1.4.0。其中，2D DFT 是对 512 × 512 的图像进行快速傅里叶变换；Resize 是将 512 × 512 的 8 比特 3 通道图像通过双线性插值运算调整为 384 × 384 的图像；Optical Flow 是用 41 × 41 窗口在 4 级图像金字塔上跟踪 520 个点构成的目标；Neural Net 用的是 FANN（Fast Artificial Neural Network Library）中的一个神经网络。

2. OpenCV 的功能

参照中文官方网站[1]，下面列出了 OpenCV 能够实现的众多功能：

1）对图像和视频数据的操作，支持文件或摄像头作为输入，图像和视频文件作为输出，进行内存分配与释放，图像复制、设定和转换数据。

2）对矩阵和向量数据的操作以及线性代数运算，包括矩阵乘积、矩阵方程求解、求取特征值以及奇异值分解等。

3）对多种动态数据结构进行操作，如链表、队列、集合、树和图等。

4）基本的数字图像处理，如可以进行图像去噪、边缘检测、角点检测、采样与插值、色彩变换、形态学处理、直方图分析和构建图像金字塔结构等。

5）对各种结构进行分析，包括连通分支、轮廓处理、距离转换、图像矩计算、模板匹配、霍夫变换、多项式逼近、线性拟合、椭圆拟合和 Delaunay 三角划分等。

6）摄像头定标，包括发现和跟踪指定模式、参数标定、齐次矩阵估计、单应矩阵估计、立体视觉匹配等。

7）运动分析，如对光流、动作分割和目标跟踪的分析。

8）目标识别，比如通过特征方法或隐马尔可夫模型（Hidden Markov Model，HMM）等。

9）基本的 GUI（Graphical User Interface，用户图形界面）功能，如图像或视频的显示，键盘、鼠标以及滚动条事件处理等。

10）图像标注，如对直线、曲线和多边形进行标注，还可以进行文本标注（目前只支持中文）。

3. OpenCV 的模块

到 2011 年 8 月为止，OpenCV 的最新版本是 2.3.1，主要包含了五个模块：

1）CV——核心函数库。

2）CVAUX——辅助（实验性的）函数库。

3）CXCORE——数据结构与线性代数库。

4）HIGHGUI——图像界面函数库。

5）ML——机器学习函数库（实现模式分类和回归分析等功能）。

在早期版本中曾出现过 CVCAM 模块，它负责读取摄像头数据，当 HIGHGUI 模块中加入 Direct Show 支持后，此模块被废除。

---

[1] http：//www.opencv.org.cn。

可以看出，OpenCV 是一个扩充性很好的算法库，其模块可以自由添加和删除，功能也在不断丰富中。因此在进行 OpenCV 编程的时候，要能不断接受新思想、新方法（事实上这正是其开源的目的和意义），经常访问 OpenCV 相关网站和论坛也不失为一种及时把握 OpenCV 最新内容的便捷途径。

1.6　主要难点与发展趋势

数字图像具有信息量大、内容丰富、表现力强、便于存储和传输等优点，在社会生活的诸多方面发挥着重要作用。但是受到计算理论和方法的制约，现有技术难以满足人们日益增长的对图像识别的广度和深度的需求，数字图像应用的突出挑战问题如下。

首先，数字图像蕴含了丰富的语义，由于图像目标固有的复杂性，出现了“有信息，用不了”的情况。在图像处理过程中，通常可方便地从图像目标中提取各种底层描述，然而，底层描述与丰富的高层语义之间缺乏简单、明确的对应关系，提取多类别、多层次的语义信息仍然十分困难。

其次，由于图像的数据量极大，出现了“信息多，用不好”的情况。随着图像数据获取手段的快速发展，图像数据量呈爆炸式增长，当前图像识别的计算模型和方法在处理高维多模态海量数据时面临着重大挑战。

最后，获取的同一组数据可用于多种用途，出现了“需求多，顾不到”的情况。数字图像的应用需求日益多样化，各种应用对处于不同概念级上的语义需求各异，同一应用在不同上下文环境下对语义的要求也不尽相同。但是，由于缺乏对需求的感知——缺乏将高层需求转化为机器可接受的高层语义特征，缺乏从高层语义特征到底层特征间可逐级计算的多层次特征表示和计算模型，因此难以从图像数据中提取出有效的语义并组织起来满足多样化的应用需求。

图像目标识别所面临的许多难点都可以归结到图像处理与模式识别领域的一些基础性问题，这些问题目前还没有满意的解决方法，但对目标识别来说它们又是如此的重要。因此，相关领域的科研人员对它们开展了大量的研究工作，并取得了一定的前沿性成果。

1. 图像中感兴趣物体的分割

大多数模式识别问题假设模式是与背景信号和其他模式分离的，目标识别也同样需要将对象从图像中分割出来，以便进一步处理。实际上，由于图像结构及其内部特征的复杂性、多样性，仅依据诸如图像颜色、梯度、纹理等底层的

图像特征很难获得反映正确对象和背景区域的分割结果，不得不借助于更高层的先验知识，而更高层的先验知识的获取和表示本身就是一个非常困难的问题。所以，尽管人们长期以来为研究对象分割问题做出了很大努力，但还是没有一种统一的理论或通用的方法能对任何情况下的任意对象进行效果理想的分割。

近年来，局部特征在目标识别中的广泛应用开辟了一条新的途径[32,33]，即不需要先将对象完整的分割出来，只是依靠检测到的对象的局部特征就可以达到识别对象的目的。随着各种局部特征不断涌现，一些研究者考虑针对不同目标的具体特性自动选择不同的局部特征来完成识别任务[34,35]，还有一些研究者正在探索将不同的局部特征结合起来进行目标识别[36,37]。

2. 视点不同造成的表象差异

在同一个场景中，视点的变化往往使得物体所呈现的表象有所不同，比如物体的大小比例、几何形状、物体的不同侧面等，这些都需要进行复杂的图像处理。对于视点远近变化造成的物体大小不同，要求识别系统具有某种尺度不变性，虽然通过多分辨率分析技术[38-41]可以部分解决这个问题，但是如何让计算机自动确定相应的尺度来识别物体，目前还没有一个令人满意的答案。

由于视点发生变化，同一个物体所呈现的不同侧面往往特征不同，甚至产生了自身的遮挡。近年来，利用三维模型建立视面图的方法[40,42,43]取得了一些成功。这种方法先是建立以三维目标为中心且与视点无关的3D模型，然后对视点进行限制并对目标进行平行投影得到二维视面模型，将目标可见表面相同的投影合并得到一个视区。针对不同视区可以提取目标在不同姿态下的特征，这种方法可以较好地解决目标姿态变化造成的目标难以识别的问题。

3. 无标记图像的学习

大部分目标识别系统经过训练后就固定不变了，或者使用相当长一段时间才重新训练一次。而实际应用时，最初的训练集中，图像的数量和代表性总是不够的，这就希望识别系统能不断地适应新的样本而不损失对原来训练过的样本的分类性能。这样的增量学习问题很早就受到关注[44,45]，提出了很多具体的方法，但还没有一个统一的理论框架。新增加的样本可能是没有类别标记的，因为无标记的图像很容易得到，而标记过程费时费力。这种同时对标记样本和无标记样本进行学习的过程称为半监督学习[46-49]，是近年来机器学习领域的一个研究热点。

目前，还有一种广受关注的目标识别问题，待识别的对象是没有分割过的图像，训练图像的标记是其中是否存在某一类物体，而不是物体的具体位置、

大小和方向。对这种标记不足的样本进行训练和识别的方法可以统称为弱监督学习[50,51]，可以用于图像检索、图像分类和目标识别等。

4. 特征推理

机器学习往往通过对样本的学习建立初始模型，采用相似性聚类或决策树等方法对个体进行预测或描述。与机器学习不同，特征推理作为解决“语义鸿沟”的另外一种途径，它是由一个或几个已知的判断（前提），推导出一个未知结论的思维过程。它不需要学习建模或者训练过程，而是实时、在线地根据已有的一些目标的特征、知识对要识别的目标进行推理和判断[10]。例如，在视频或连续图像中进行运动目标检测，如果我们能识别出当前场景是一条马路，并且知道轮船只能在水上行驶，就可以推断出马路上发现的运动目标一定不是轮船。

视觉认知的整体性和层次性决定了整体特征和局部特征的存在，推理往往基于不同尺度的特征进行归纳、演绎，完成推导过程。如图 1-4 所示，在基于多源卫星影像的目标识别系统中，首先在粗尺度上提取特征，对于多光谱和合

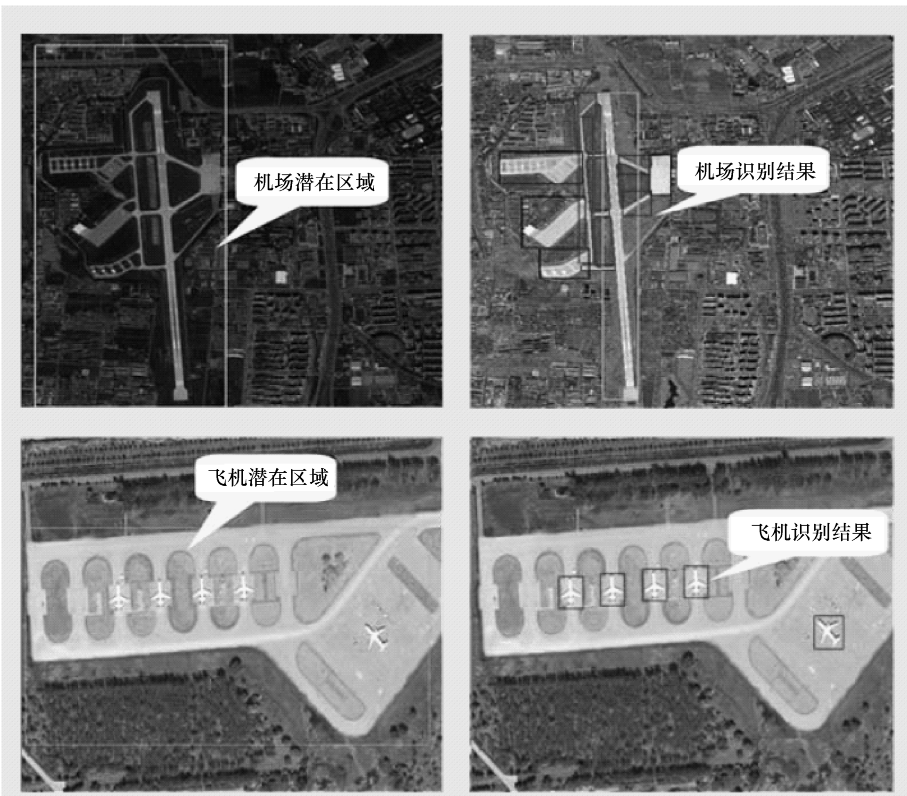


图 1-4　通过特征推理进行图像目标识别的示例（来源：李波，2011 年）

成孔径雷达影像，根据光谱或散射特性把影像分割为水体、陆地、人工建筑、植被等，根据目标的环境特性给出目标可能出现的区域，缩小目标搜索范围；进而对多光谱、全色和遥感影像提取纹理、形状、边缘等特征，综合整体特征、局部特征和知识库，在机场潜在区域搜索飞机，通过面向对象的推理实现飞机的识别。

1.7 研究内容与结构安排

本书主要针对可见光图像和刚性目标，学习并借鉴了图像工程、模式识别、机器视觉和人工智能学科中一些先进技术，探讨了复杂背景下的目标识别以及局部遮挡物体的识别中的关键问题，为增强现有图像识别系统的自动化程度和信息处理能力提供理论支持和技术帮助。

1.7.1 本书的研究内容

本书围绕着图像目标的表示与识别这一主题，鉴于当前国内外相关领域的众多先进成果和空白之处，对以下几个方面的问题进行了深入的探讨和研究。

1. 特征提取技术

目标特征提取是目标识别中的关键技术，对于识别的最终效果有着决定性的影响。整体特征和局部特征各有自己的适用范围，都要求对亮度、尺度、平移和旋转具有一定的不变性，从广义上讲，它们的提取过程都包括特征生成和特征优化。其中整体特征的性能取决于目标分割的准确程度，局部特征的性能在很大程度上取决于特征区域的选取和描述。本书根据应用背景和实际需求，详细阐述了整体特征的提取过程和相关技术，根据应用背景和实际需求，选用并改进了一些特征区域检测算法和特征区域描述算子，为目标匹配和分类提供了性能优良的局部特征。

2. 目标匹配技术

使用模型直接匹配未知物体，并选择最佳匹配为最终识别结果，是在很难得到有关特征概率和类别概率的先验知识，或者得到的数据不足以设计分类器的情况下的目标识别方法。而图像背景复杂度、图像清晰度、图像中目标数目和局部遮挡等因素对图像目标匹配识别的效率、可伸缩性和适用性提出了挑战。本书对匹配方式和相似度度量的研究现状进行了深入分析，仔细研究了通过局

部特征进行目标匹配的相关算法，针对目标匹配在图像拼接和图像检索中应用的不足之处，提出了基于多分辨率技术的航拍图像拼接方法，以及基于原型匹配的图像检索方法。

3. 目标分类技术

目标分类一般需要构造有效的特征向量和充分利用相关领域的知识，而且设计分类器是目标分类的主要任务和核心研究内容之一。本书详细介绍和比较了几种典型的图像目标分类器的原理与特点，并综合评述了分类器的不同种类以及性能评估方法。向量空间模型最初是模式识别领域中常用的文本表示方法。由于局部特征性能优越，含有的局部信息可以对图像的内容进行多语义层次的描述，也为利用向量空间模型进行目标表示提供了一条有效途径。本书针对当前局部特征在目标分类中应用的不足之处，充分借鉴了向量空间模型的思想，并结合信息论的相关技术进行特征优化，提出了一种基于局部特征的目标分类方法，在标准图像库上的实验结果证明了该方法的有效性和鲁棒性。

4. 视点变化下的目标识别技术

视点变化造成目标的表象差异是目标识别领域的一个难点，尤其是观察角度发生变化，同一物体的不同侧面呈现出迥异的特征，甚至产生了自身遮挡的问题（物体的某个部分遮挡了该物体的其他部分）。本书通过对三维物体进行视面模型表示，得到了目标不同姿态的二维投影描述，从而为视点变化下的目标识别构建了合适的模型库。通过对角点特征的深入研究，结合主分量法和Hausdorff距离，提出了一种在视点变化下目标匹配识别方法；并提出了基于角点标记图的 BP 网络分类方法。实验对比证明，基于该特征的识别算法在视点发生变化时对目标的识别更为有效。

1.7.2　本书的结构安排

本书的组织结构如下：

第 1 章，绪论。介绍了本书的研究目的和意义，并给出了图像目标识别的定义、系统框架和两种研究思路；列举了图像目标识别常用的图像库，探讨了图像目标识别的主要难点和发展趋势；最后，对本书基本内容和结构安排进行简要说明。

第 2 章，图像目标的整体特征提取。讨论了图像分割和目标分割的关系，介绍了目标分割的研究现状和基本方法；利用三类整体特征对目标进行表示与

描述；分析了目前主流的特征空间优化技术。

第 3 章，基于整体特征的目标识别。概述了模式识别的基础理论和方法；对目标匹配技术和目标分类技术的研究现状进行了回顾；讨论了目标匹配的两种基本方法和四种基于距离的相似度度量；详细论述了常用的图像目标分类器的设计和训练方法。

第 4 章，图像目标的局部特征提取。讨论了局部特征的含义和局部特征提取的通用步骤方法；在 DoG 特征点检测的基础上结合 SIFT 和 GLOH 描述子完成了对复杂图像的局部特征提取与描述；在狭义特征点——角点的检测技术研究中，针对 SUSAN 算子固定阈值的问题，提出了自适应阈值的改进方法。

第 5 章，基于局部特征的目标匹配。提出了基于最邻近距离比（NNDR）与霍夫变换的特征匹配策略；针对局部特征匹配在目标图像拼接和图像检索中应用的不足，提出了基于多分辨率技术的航拍图像拼接方法，以及基于原型匹配的图像检索方法。

第 6 章，基于局部特征的目标分类。详细介绍了目标的向量空间模型表示；阐述了视觉单词的理论依据以及基于 RNN 算法的视觉单词库特征库构造方法；在此基础上，结合信息论的相关技术进行特征选择，提出了一种基于局部特征的目标分类方法。

第 7 章，基于角点特征与视面模型的目标识别。通过三维物体的视面模型表示方法构造目标在不同姿态下的投影模型库；利用基准角点定义了一种具有平移、旋转、尺度不变性描述子并用以识别飞机目标；结合主分量法和Hausdorff 距离，提出了一种在视点变化下目标匹配识别方法；提出了基于角点标记图的 BP 网络分类方法。

第 2 章　图像目标的整体特征提取

科学家必须在庞杂的经验事实中抓住某些可用精密
公式来表示的普遍特征，由此探求自然界的普遍原理。
——阿尔伯特·爱因斯坦（1879—1955）

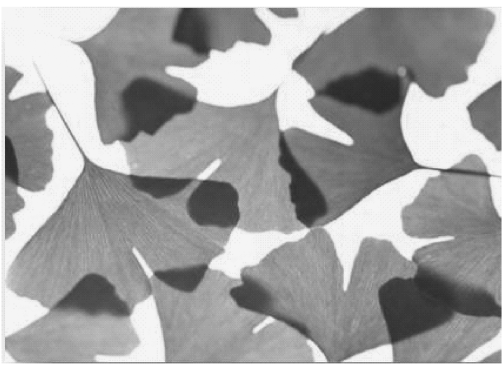
2.1　引言

认知科学上关于视觉的相关理论认为，特征是决定相似性与分类效果的关键，当分类的目的决定之后，如何找到合适的特征就成为认知与识别的核心问题。目标识别系统通常要提取具有如下性质的特征描述：来自同一类别的不同样本的特征值应该非常相近，而来自不同类别的样本的特征值应该有很大的差异。这样我们就产生了提取最有“鉴别（Distinguishing）”能力的特征的想法，这些特征对与类别信息不相关的变换具有不变性（Invariant）[52]。

这种抓住本质特征来表示目标的方法，一般称之为模型方法，而物质世界的统一性是模型方法的哲学基础。自然界和社会生活中的各种各样的事物，都是运动着的物质的各种不同的形式，在千差万别中存在着同一性，如外形结构相似，生理的、心理的过程相似，物理过程相似，功能、行为相似，以及不同运动形式可以用共同的数学方程式来描写等[53]。

图 2-1a 中所示为各式各样的银杏叶，每片叶子都体现着事物的特殊性，这也正印证着德国哲学家莱布尼茨的名言，“世界上没有完全相同的两片树叶”，如果不能针对事物的普遍性，只是一片一片地观察所有的个体，即使“认识银杏叶”这种非常简单的问题也无法表述，更无法解决。图 2-1b 所示的这个银杏叶的模型，就是抽象思维和形象思维统一的一个例证，一方面，它抽象出被研究对象——银杏叶的形态本质，避免了对每个个别事例进行全面描述的繁琐过程；另一方面，它运用形象思维的手段（如图形、符号等）来反映事物的形态本质，具有直观性、鲜明性和生动性。在科学探索中，模型（用特征来表示目标）在这两方面都发挥了重要的认识论功能。

从方法论的角度来看，建立模型的关键是要从错综复杂的矛盾中抓住主要矛盾，要在尽可能周密地进行具体分析的基础上舍末求本，撇开次要的因素、关系和过程，突出主要的因素、关系和过程，找到对事物的发展起决定性影响的因素和规律。在图像目标的表示和识别过程中，提取特征建立模型既要照顾到真实性，特征大体要能反映出目标的主要方面（光谱、纹理、形状等），又要做到尽可能简化，使得建立的模型是当时已经掌握的理论工具和数学方法所能处理的问题。



a)



b)

图 2-1　银杏叶模型的特征示例

a）银杏叶图像　b）银杏叶模型

基于上述性质可以看出，提取特征建立模型的过程中要求遵守一定的方法论原则。

1. 相似性与简单性的统一

从相似性来说，不可能也不必要要求模型和目标本身在外部形态、质料、

结构、功能等所有方面完全一致。但是必须按照所研究问题的性质和目的，使模型与目标本身具有本质上的相似。从简单性来说，就是要化繁为简，化难为易，使复杂物体有可能通过比较简单的模型来进行研究。模型具有简单性才能够实行操作，真正发挥作用。但简化不是主观随意的，必须以不丧失模型与目标本身的本质上的相似度为原则。在用模型逐步逼近目标本身的过程中，既要保证模型应有的精度，又要尽量合理简化，坚持两者的统一。

如图 2-2 中，毕加索最后抽象出来的公牛的特征，排除其在艺术上的加工，基本上抓住了公牛的本质特征，坚持了相似性与简单性的统一，在绘画的角度上使模型与目标本身具有本质上的相似，又从公牛各种各样复杂的表象中合理简化，使得这个模型非常容易和其他动物的模型相区分。

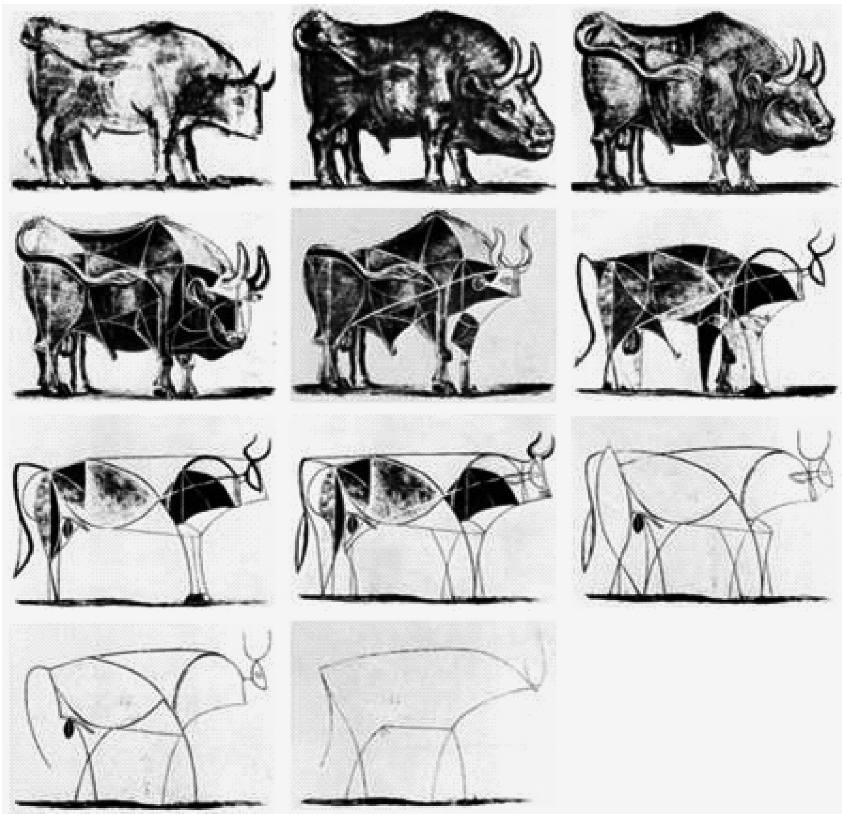


图 2-2　**毕加索的公牛**

2. 具有可验证性

一般来说，只要模型具有可操作性，就有具体的操作过程，并能够取得具

体的研究结果，这结果是可以实际进行对照和比较的，因而就是可验证的。如果通过检验发现了模型的缺陷，必须对模型加以修改，甚至代之以新的模型。如果模型经受了检验，也还需要进一步从理论上论证其科学性。

从某种角度上来说，牛的牛角、牛尾、四肢是非常显著的特征，以此建立的模型从绘画角度上具有很强的可操作性。当然，在图像目标的表示与识别中还要具体问题具体分析，比如区分牛和飞禽、房屋等，这些特征同样还是适用的，但是如果要细分出几种牛，或者区分出公牛和母牛，这些特征恐怕已经难以经受实践检验了，必须对特征加以修改，或者提取新的特征。

3. 多种知识和方法的综合运用

无论是建立模型，还是运用和检验模型，都没有刻板的程序和完全固定的方法。一个有效合理的科学模型，既要严格以目标本身为依据，又要求人们广开思路，使经验方法和理论结合，逻辑思维和非逻辑思维并用。模型的综合性的特点，决定了建立模型需要综合地灵活地运用多种多样的思想、知识和方法，充分发挥自己的形象思维能力。

针对图像目标识别的性质和目的，要综合运用多种知识和方法，不能以偏概全，仅仅以一种特征建立模型来进行各种识别任务。毕竟，把两种目标区分开来，把几种目标两两区分开来，还有把一种目标和其他所有目标区分开来，复杂度可能远远不在一个级别上，需要的知识和方法可能也不在一个层次之上。而识别目标个体与识别目标类别也有很大的不同，需要经验方法和理论方法的结合，加以灵活运用。

在模式识别领域，从狭义上讲，特征提取就是特征形成，即根据被识别的对象产生出的一组基本特征，它可以是计算出来的（当识别目标是波形或数字图像时），也可以是用仪表或传感器测量出来的（当识别目标是实物或某种过程时），这样产生的特征叫做原始特征[54]。从广义上讲，特征提取还包括特征空间的进一步优化。

根据特征描述的区域范围不同，图像目标特征又可以分为整体特征和局部特征两个大类。整体特征是针对已经分割出来的目标而言的，对一个图像目标整体进行特征表示，进而分类决策。整体特征在图像目标的表示与识别中的应用瓶颈主要是由于其提取效果依赖于目标分割的准确度，而目标分割本身就是一个复杂的工作，分割过程中出现的任何误差都有可能影响到后续的目标描述与分析。

2.2　图像目标分割

图像目标分割的主要工作是将图像中特定的目标对象与背景图像进行分离，它在数字图像处理与计算机视觉领域应用越来越广泛。例如，在执行交通监控的车牌识别时，常需先从整个道路场景图像中分离包含待识别车牌的区域；在图像编辑过程中，使用频率最高的一组操作便是将一幅图像中的某个兴趣对象复制出来，并将其粘贴到另一幅背景图像中合成一幅新图像；在医学图像处理中，将脑部磁共振成像（Magnetic Resonance Imaging，MRI）图分割成脑组织（包括灰质、白质、脑脊髓等）和非脑组织区域，然后在此基础上进行配准、三维模型重建等高层处理；在基于内容的图像压缩、检索等应用中，将图像分割成具有不同物理意义的目标区域，然后针对不同的目标采用合适的方法，以实现更高效的压缩和检索。

2.2.1　图像目标分割概述

对整体特征进行有效的理解和研究，必须要明确目标分割的定义以及它与图像分割之间的关系。图像分割指将图像划分成若干彼此互不交叠且自身具有某种相似属性的同质区域。通常它包含较广的含义，进一步可以细分为面向图像特征的图像分割和面向物理、语义特征的目标分割。其中，狭义的图像分割主要强调图像的区域和边缘，力求区域间的特征差异较大，而区域内差异最小，其分割结果将形成互不交叠的图像区域或者轮廓线；而目标分割特指将具有物理、语义特征的目标对象从相应的图像背景中分割出来，强调两者分离，其中，背景可以是其他单独的对象或其他任意对象的集合，而目标对象，通常又称之为前景，则是图像中客观存在的具有某种物理或语义意义的实体。令 I 表示一幅 $n\times m$ 的待分割图像，则图像分割的定义可形式化的表示为将 I 划分为满足下述条件的 N 个子区域 I_i，$(i=1,2,\cdots,N)$[55,56]：

1）$\bigcup_{i=1}^{N} I_i = I$；

2）$I_i \cap I_j = \varnothing$，其中 i，$j=1,2,\cdots,N$ 且 $i\neq j$；

3）$S(I_i)=\text{TRUE}$ 且 $S(I_i\cup I_j)=\text{FALSE}$，其中 i，$j=1,2,\cdots,N$ 且 $i\neq j$，$S(I_i)$是对 I_i 中所有元素属性相似性描述的逻辑谓词。

其中，条件 1）指出图像分割的结果需满足该图像可由分割产生的所有子区域组合而成；条件 2）指出图像分割结果中的任意两个子区域不存在公共元素，

两两互相不重叠；条件3）指出图像分割结果中每个子区域内部相似属于同质区域，而子区域之间则有差异，或者说属于同一子区域的元素具有一些相同的特征，而属于不同子区域的元素的特征不同。

这里，图像分割与目标分割均可以采用上述形式化定义，不同的是图像分割的逻辑谓词 $S(\cdot)$采用的是图像颜色（包括灰度和彩色）、纹理、梯度等图像低层特征，其分割结果与实际物理对象之间并不一定存在一一对应关系，而目标分割利用了更高级、抽象的对象特征，强调分割结果中目标对象与背景的分离。

此外，在第1章关于目标识别系统三个层次的计算处理（1.3.2节）中，狭义的图像分割仅仅属于低、中层视觉问题，其处理过程仅依赖于原始数据本身，虽然可以使用极少量的先验知识（如预先设定的阈值等），但却不依赖于这些先验知识；随着视觉层次的提升，所能利用的先验知识也越来越丰富，对先验知识的依赖程度也越来越高，目标分割问题属于中、高层视觉问题，可以借助目标对象的外观、形态、轮廓等高层先验知识来实现对目标对象的分割。图2-3所示为图像分割与目标分割的例子。

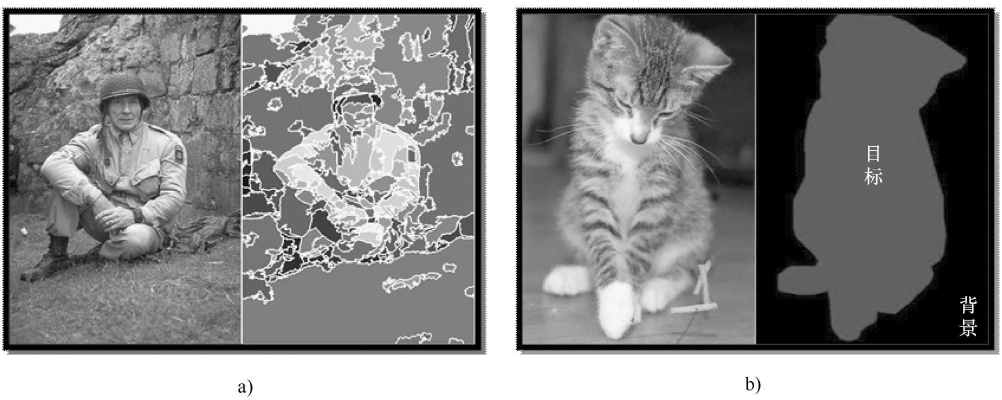


图2-3 图像分割与目标分割（来源：刘陈，2009年）

a）图像分割 b）目标分割

2.2.2 图像目标分割现状

早期的图像分割研究主要集中于对狭义图像分割的研究，且并未区分狭义图像分割与目标分割的概念，界定比较模糊。随着数字图像处理和计算机视觉研究和应用的不断发展，更多的需求强调针对图像中某些具有特定物理、语义

意义的兴趣目标的分析处理，使得目标分割技术得到了更广泛的关注和研究。但是，由于图像结构及其内部特征的复杂性、多样性，仅依据诸如图像颜色、梯度、纹理等原始图像特征很难获得反映正确目标和背景区域的分割结果，不得不借助于更高层的先验知识。这些高层先验知识主要是人们对于待分割目标的认识和理解，并通过形式化的方法加入到分割过程，从而使得分割方法能将目标对象与背景分离。

然而，至今仍没有一种统一的理论或通用的方法能对任何情况下的任意目标进行理想的分割，甚至在同一种情况下，都做不到所有方法都能获得好的分割结果。造成这种结果的原因包括客观和主观两个方面。客观原因分析如下：

1）图像获取的途径多样，成像原理、技术手段各异。常见获取数字图像的设备有各式各样的数字摄像机、照相机、扫描仪等；而成像的原理和技术更是各有不同，有激光、红外以及 X-射线、超声波、CT（Computer Tomography）、MRI（见图 2-4）等。不同的获取途径、成像原理和技术造成了图像的情况多样，质量不一。

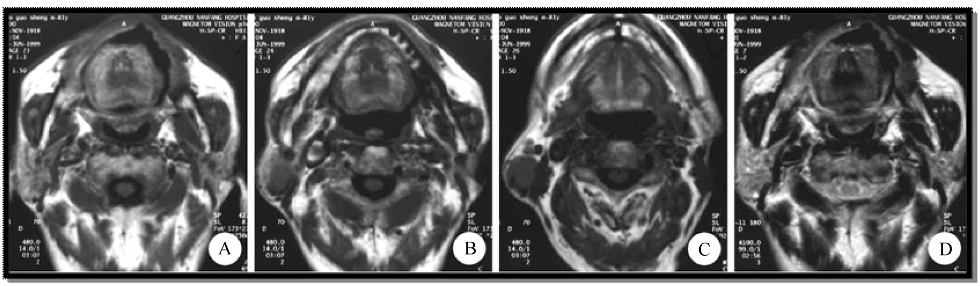


图 2-4　某右腮腺淋巴瘤 MRI 图像

2）图像本身结构复杂，内部特征多样。从颜色空间的角度来看，图像可以分为二值图像、灰度图像和彩色图像，而彩色图像又包括 RGB、HSI、YUV 空间等不下 10 种，各空间特点不一；从图像空间的角度又可分为普通图像和纹理图像，当图像区域一系列的局部特性是稳定的、缓慢变化或者近似周期的，则该图像区域具有不变的纹理，如图 2-5 所示，而且，除了对自然场景成像得到的图像之外，还有大量的艺术创作图像，如图 2-6 所示。

3）图像仅是现实世界的表象。图像仅仅是现实世界在图像平面的成像，由于成像过程中的复杂因素如光照、遮挡、3D 到 2D 的深度信息丢失等所造成图像信息的损失，图像的特征仅仅是真实特征的表象，并不能完全等同于真实目标，即真实特征的差异有时并没有明显的表象差异与之相对应，如目标对象和

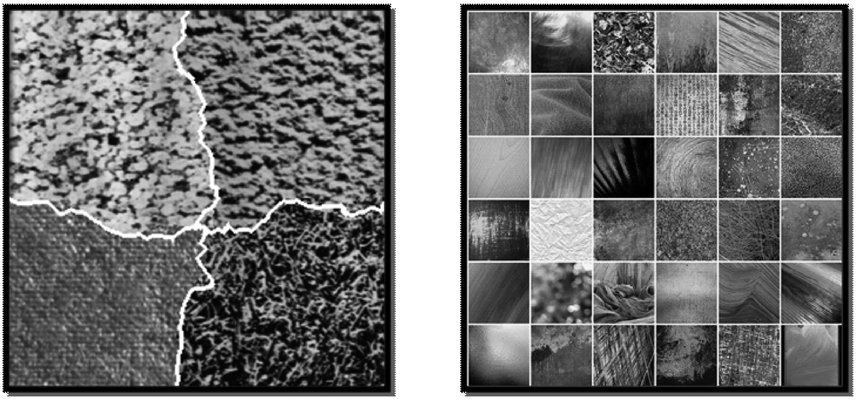


图 2-5　纹理图像图例

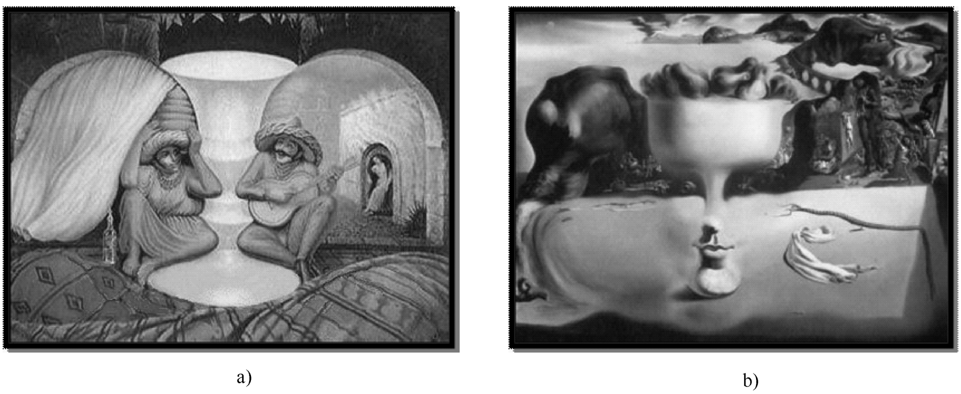


图 2-6　视觉幻象和超现实主义艺术图像（来源：Octavio Ocampo 和 Salvador Dali）

a）Forever Always　b）Apparition of Face and Fruit Dish on a Beach

背景由于对比度降低，而在边界处混淆在一起不易区分；目标对象与背景具有相似的颜色或纹理等情况。因此，很难仅根据简单的图像表象特征，如图像的颜色、纹理、边缘等获得正确的图像目标区域，而不得不借助于更高层的先验知识[57]。

主观原因分析如下：

1）建立统一的数学模型存在较大困难。由于面临的实际问题不同、分割目的不一，导致图像中的目标对象和背景并非固定不变，而是在不同的需求和应用下具有不同的定义和内容。此外，研究者们自身知识结构的局限也导致无法给出适用所有情况的统一数学模型。

2）受到相关学科发展的制约。目标分割是多学科交叉的研究领域，受到诸

如模式识别、机器学习、数值优化方法等学科的影响，虽然近年来，这些领域均取得了较明显的进步，但离使计算机具备像人脑一样复杂的分析处理能力还有很大的距离。只有随着各学科的综合发展，目标分割才会不断有新的突破。

3）无法给出一致的用户满意度标准。一方面，尚没有对目标分割方法和分割结果真正客观的评价标准，通常只能人为地建立一组有限的实验图像以及其对应的真实分割结果，然后通过实验对比分割方法在该组图像上某几方面的性能；另一方面，由于需求差异以及用户主观方面的原因，也导致即使同样的分割结果在不同的应用背景下，对不同的用户也可能存在完全不同的评价结果。

总之，尽管长期以来人们为研究目标分割问题做出了很大努力，但上述原因导致很难实现一种普适的方法，而只能针对特定问题和具体的需求给出合理的解决方法，在处理速度、精度等关键性指标上做出均衡或侧重。

2.2.3　图像目标分割技术

图像目标分割技术历经数十年的发展，其中用到的算法种类繁多、不可胜数，虽然本书将图像分割分为狭义图像分割和目标分割，但这两者中的许多概念、思想和方法都有着非常密切的联系，而且前者是后者的重要基础。因此，不能简单地抛开狭义图像分割而谈目标分割，有必要对两者进行综合的分析和论述。

近年来，涌现了许多不同的图像分割分类标准，比如，按照用户参与的程度可分为自动、交互式与纯手工的分割方法；根据利用区域内相似性还是区域间相异性原理的区别可分为基于区域、基于边界或者两者结合的算法；依据分割结果的确定性与否可以分为软分割与硬分割等。这些划分都比较粗糙，不能很好地体现狭义图像分割与目标分割各自的特点，本节根据狭义图像分割与目标分割最显著的特点，即狭义图像分割一般通过数据驱动，而目标分割往往需要知识驱动，对两者展开介绍。

1. 数据驱动的图像分割

图像分割算法一般基于亮度值的两个基本特性之一：不连续性和相似性[8]。针对第 1 个特性，可以利用亮度的不连续变化分割图像，如图像的边缘。针对第 2 个特性，可以依据事先制定的准则将图像分割为相似的区域，门限处理、区域生长、区域分类和聚合都是这类方法的实例。围绕着这两个基本特性，传统的图像分割方法又可以粗分为基于边缘的分割、基于阈值的分割、基于区域的分割三个大类。

（1）基于边缘的分割方法

边缘检测是在灰度图像分割中广泛应用的一种技术，它基于在区域边缘处梯度变化剧烈的假设，试图通过检测区域间的边缘来达到图像分割的目的。在灰度图像中，梯度由相邻像素的灰度级差异表示，常用的灰度图像边缘检测算子有 Sobel 算子、Laplacian 算子、Laplacian of Gaussian（LOG）算子、Canny 算子等。

根据数学特性又将这些算子分为两类：与 Sobel 算子类似的称为一阶微分算子，而 Laplacian 算子、LOG 算子、Canny 算子均属于二阶微分算子。一阶微分算子利用的是图像在 X 或 Y 方向上的一阶导数在边缘处取极值或 0 的特性，而二阶微分算子则利用 X 和 Y 的二阶导数。在彩色图像中，边缘来自于三维颜色空间的突变，可以将现有的灰度边缘检测技术直接应用于彩色图像的每个分量，再根据一定的方法进行合并，常用的合并方法有均方根、求和、取最大绝对值等。

除了直接利用边缘检测算子提取图像边缘外，还有一些方法也相继被提出，如边缘松弛法、边界跟踪、图像滤波、多尺度变换和主动轮廓（Active Contour）等。如图 2-7 所示，基于边缘检测的方法仅利用了图像的梯度信息，当图像质量较好时定位精度高，但受噪声和图像质量的影响常常会检测出伪边缘，导致错误的分割结果。而且一组边缘像素点很少能完整地描绘目标的轮廓，因此，典型的做法是在使用边缘检测算法后紧跟着使用连接过程，将边缘像素组合成有意义的边界。



a)　　　　b)

图 2-7　对图像进行边缘检测

a）原图像　b）分割结果

（2）基于阈值的分割方法

阈值法是一种最为简单的利用颜色信息进行图像分割的方法，在灰度图像分割中，它基于这样的假设：同一区域的内部像素，它们的灰度值相似，但不同区域的像素灰度差异较大，其在灰度直方图上的反映就是不同的区域对应不同的波峰。则分割时，选取的阈值应位于直方图波谷处。按照选取域值的数量又可分为单阈值法和多阈值法。在单阈值分割中，分割的结果为两类区域；在多阈值分割中，分割的结果为多类区域。

对于彩色图像而言，由于其包含 3 个颜色分量，在三维直方图中确定阈值是比较困难的，如果阈值的选择分别在每个颜色分量上单独进行，则忽略了 3 个分量间的相关性，导致分割结果不准确。通常，彩色图像的阈值分割采用降维的方法，从三维颜色空间向低维投影，形成二维平面或一维直线，然后在低维上选择合适的阈值[57]。

基于阈值的分割方法实现简单，但存在以下明显的缺点：对于不存在明显波峰和波谷的直方图（受噪声干扰，或者彩色图像各分量的直方图本身就可能不存在明显波谷），得不到满意的分割结果；仅考虑了图像的颜色（灰度）信息，而忽略了图像的空间信息，所以对噪声非常敏感，如图 2-8 所示，岸边的白色建筑物也被当做船体上的目标了。



a)　　　　b)

图 2-8　对图像进行二值化分割

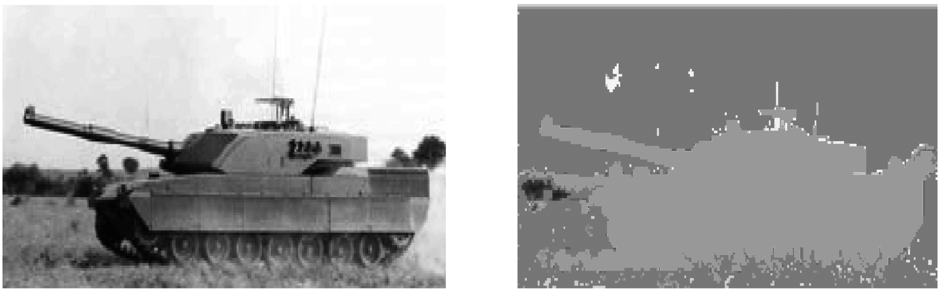
a）原图像　b）分割结果

（3）基于区域的分割方法

基于区域的图像分割考虑了图像的空间信息，如图像灰度、纹理、颜色和像素统计特性等，进而将目标对象划分为同一区域的分割方法。常见的区域分

割方法主要有：区域生长、区域分裂、区域合并和分水岭分割方法。

区域生长和区域分裂是两种典型的串行区域技术，区域生长法的基本思想是：根据一定的相似性原则，将满足这一原则的像素合并起来构成区域，其关键点是生长种子和生长准则的选取，效果如图 2-9 所示；而区域分裂法恰恰相反，则是将整幅图像作为原始分割结果，当分割结果不能满足一定的均匀、相似性时，就将其分裂，直到每个区域内部都相似为止。



a)　　　　b)

图 2-9　对图像进行定点区域生长

a）原图像　b）分割结果

两者结合的方法通常又称区域合并，它将相邻且具有相似的区域合并，而将明显不相似的区域进行分裂。基于区域生长、区域分裂的方法受噪声的影响比较小，效果优于阈值法，但区域生长依赖于种子点的选择和生长顺序，而区域分裂则可能会使边界被破坏。

分水岭分割方法，是一种基于拓扑理论的数学形态学的分割方法，基本思想是将图像看做测地学上的拓扑地貌，像素的灰度值表示该点的海拔，每一个局部极小值及其影响区域称为集水盆，而集水盆的边界则形成分水岭。分水岭的概念和形成可以通过模拟浸入过程来说明。缺陷：对噪声极为敏感，易产生过分割现象；相关研究人员[8]提出了一种将分水岭算法与自动种子区域生长相结合的分割算法，有效解决了算法中过分割的现象，如图 2-10 所示。

2. 知识驱动的目标分割

主动视觉理论[14]的建立，为利用高层先验知识指导目标分割提供了强有力的方法和理论依据。近年来，越来越多的高层先验知识在计算机上通过各种形式表达出来，并且与低层图像特征相结合，共同指导图像目标分割。本节从分割过程中所采用的先验知识的种类出发，对各种目标分割方法进行分类阐述（目标分割往往采用多种先验知识，而不是单独一种，这里不作严格的区分）。



a)　　b)

图 2-10　对图像进行分水岭分割

a）原图像　b）分割结果

（1）外观信息

外观信息是对某一个或某一类目标（如人、汽车、草地、海、天空）外表特性或共性的描述，包括外观颜色、纹理等。典型的用于表示外观信息的方式有种子点、统计直方图、聚类、有限混合模型等，通常外观模型需要根据一定数量的样本数据，经过机器学习的方法训练而来。在应用外观信息进行目标分割的方法中，最简单的是 Adobe 公司的图像处理软件 Photoshop 中的魔棒工具（Magic Wand），它通过在一定容差范围内，寻找与用户指定的种子点相匹配的像素，完成目标分割，但由于通常情况下目标对象外观并非简单的由某几个种子点就能正确表达，而且它孤立的考虑颜色的匹配度而没有考虑到像素的空间相关性，所以分割效果常不能令人满意。

Wang 等人[58]以局部的目标对象和背景外观样本像素为起始点，利用信念传播方法不断地对周围像素进行前、背景估计，并同步更新外观颜色模型，随后又提出校验用于更新外观颜色模型的样本的方法，并成功地实现了实时的目标分割工具；Growcut 利用细胞自动机的原理，以用户输入的目标和背景样本为起始点，迭代地对其外围像素进行“竞争蚕食”，最终实现稳定的“群落”而完成分割；随机游走计算其他像素随机游走到达各样本笔划的概率，取概率最高的笔划标签作为对像素的分割；Boykov 等人[59]利用灰度统计直方图作为外观模型，同时结合图像对比信息（利用相邻像素灰度级的 L_2 范数计算得来）进行图

像和视频目标分割，并在其基础上结合基于 level set 的配准目标轮廓作为形状先验进行医学图像目标分割。在彩色图像分割中，大多采用高斯混合模型代替直方图模型来作为外观颜色模型。

（2）形变信息

形变信息通常由形变模型指定，形变模型一般指主动轮廓模型（Active Contour Model），是基于微分几何、弹性力学等数学和物理工具定义的一类具有变形能力的模型。由于主动轮廓模型一般是基于目标轮廓的正则化约束（如连续性、光滑性、封闭性等），而不是目标形状信息，所以也被称为自由形变模型（Free-form Deformable Model）。它又可分为参数化主动轮廓模型（Parametric Active Contour Model，PACM）和几何主动轮廓模型（Geometric Active Contour Model，GACM）。

最早的 PACM 由 Kass 等人于 1987 年提出，通常又称它为 Snake 模型。其原理就是使轮廓模型在外力和内力的作用下向目标的边界逼近，外力推动轮廓曲线向边界移动，而内力保持轮廓的光滑性。在数学上，轮廓可表示为参数曲线：

$$C(s) = [x(s),\ y(s)],\ s \in [0,\ 1] \tag{2-1}$$

而最终需寻找使下式中内能和外能加权最小的参数化曲线：

$$E(C) = \int_0^1 [E_{\text{int}}(C) + E_{\text{ext}}(C)]\mathrm{d}s \tag{2-2}$$

内部能量表示为

$$E_{\text{int}}(C) = \frac{1}{2}\left[\alpha\left|\frac{\partial C}{\partial s}\right|^2 + \beta\left|\frac{\partial^2 C}{\partial s^2}\right|^2\right] \tag{2-3}$$

式中，α 和 β 分别表示曲线的弹性和刚性系数。一阶项保证曲线被均匀且不过度拉伸，二阶项用来减小曲线的曲率。

外部能量表示为

$$E_{\text{ext}}(C) = -\left|\nabla[G_\sigma(C) * I(C)]\right|^2 \tag{2-4}$$

式中，I 为灰度图像；$*$ 为卷积算子；G_σ 是标准差为 σ 的二维 Gaussian 函数。

PACM 将目标对象轮廓的连续性、光滑性及封闭性等先验约束知识与低层图像特征（这里是边缘、梯度特征）巧妙的结合，有效地解决了原来目标边界提取时出现病态、没有唯一解的情况。但是，它也存在以下不足：模型初始化需人工参与且对初始位置较为敏感；曲线参数化后精度不高；求解能量时容易陷入局部极值；不具备拓扑结构自动变化能力（曲线分裂、合并），不能同时分割多个目标对象；外力场的作用范围小等。

针对参数化轮廓模型的不足，相关领域的研究人员主要从以下几方面对它

进行了改进：在轮廓线方面，提出了基于 B-样条的 PACM，基于 NURBS 的 PACM 和用 Fourier 级数表示轮廓线的 PACM；在轮廓线拓扑变化能力方面，T-PACM通过在图像区域建立三角网格解决了拓扑变化的问题；在求解能量函数的方法上，提出用有限元方法、神经网络、动态规划、贪婪算法和遗传算法等替换原来的有限差分法进行优化求解；由于外力场的作用是驱动轮廓曲线向目标边界运动，对于 PACM 的性能上起着至关重要的作用，故而成为 PACM 研究的关键，其中最为著名是气球力模型（Balloon Force）和梯度向量流模型（Gradient Vector Flow，GVF）。

GACM 正在逐渐成为图像目标分割的研究热点，它是以曲线演化理论（Curve Evolution Theory）以及水平集方法（Level Set Method）为基础的活动轮廓模型。与 PACM 一样，它通过与低层图像特征结合来恢复目标对象的边界，不同的是 GACM 的轮廓线是用一个更高维水平集函数的等值曲线来隐含地表示的，通过不断更新这个水平集函数达到曲线演化的目的，而利用有效地更新水平集函数，即可随意地改变所表示曲线的拓扑，从而克服了 PACM 不能分割具有复杂边界或拓扑的目标，也不能同时分割多个目标等拓扑变化的问题。

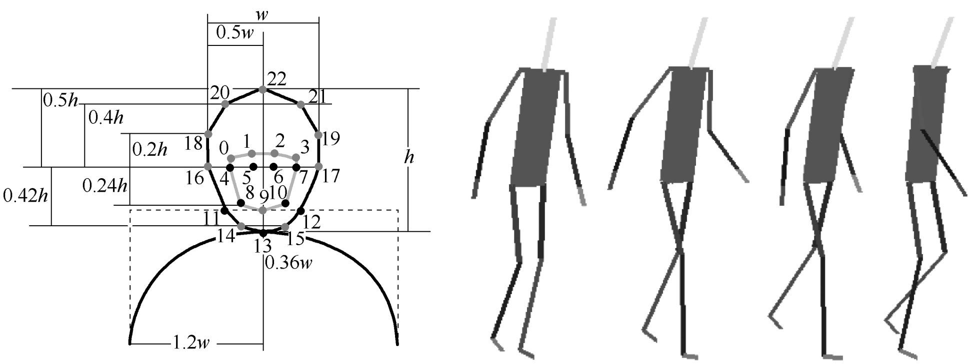
（3）形态信息

形态（包括形状和姿态）信息是比正则化约束更具体、更高层的对目标的认识和理解，通常“有形”的目标在其形态上会存在共性以及局部范围的变化，对目标形态的描述就是对这种共性和变化的描述，通常的描述方法有基于原型的方法和基于解析式的方法。解析式方法是当目标形态的几何结构比较好，即可以由一族曲线或几何图形（近似的）表示时，通过参数化的解析式的方式来定义目标形态模型。解析式描述了目标形态的共性，而参数的定义域则确定了形态的变化范围。当目标形态不能通过解析式的方法确定时，基于原型的方法提供了较为合理的解决方案，它常以二值图像模板的形式来刻画目标形态，以一组具有代表性的模板来确定目标形态的变化范围，最后，通过匹配的方法与图像特征对应。

最典型的基于原型的方法是利用模板的平移、旋转、缩放等简单变换，使其与图像特征（如边缘）匹配，最后，自动进行目标轮廓提取。然而，实际情况是，同类目标个体存在差异，而不同类目标在形态差异可能更大，简单的变换并不能有效地解决这种差异。因此，除非模板库足够庞大，能包罗万象，否则因为这些差异导致的分割不准确将在所难免。可以利用主分量分析法对模板中的形态进行学习，得到平均形态和变形参数，而其主动外观模型（Active Ap-

pearance Model）更是考虑形态模型内部的外观信息，在纹理图像分割方面表现突出。

在基于解析式对形态进行参数化描述的方法中，橡皮模型（Rubber Mask）以及画报结构模型（Pictorial Structure Model，PSM）是较早期的方法。Kumar 等人[60]利用学习而得的牛和马的 LPSM，通过各层与图像低层特征匹配，结合各层的外观信息完成对牛和马的自动分割。如图 2-11a 所示，Wang 等人[61]利用人脸检测技术结合人体头和肩的解析式形态模型实现人体上身的定位，通过迭代图割优化不断地更新外观 GMM 实现对人体上身的自动分割。如图 2-11b 所示，骨架模型（Skeleton Model）也是一种常用的表示人体或铰链模型（Articulated Model）形态的模型，Kohli 等人[62]利用它实现了多视图间目标分割与姿态估计的同步。



a)　　　　b)

图 2-11　解析式形态模型举例

a）人体上身模型　b）骨架模型

（4）目标识别引入的信息

目标识别引入的信息通常是与目标对象所属的类相关的（Class-specific），是对目标形状、轮廓、外观等特征所具有的共性的提取和学习，通过图像中的角点、线、边缘、图像块等形式给出（典型的特征检测算子与描述算子见本书第 4 章内容）。Borenstein 等人[63]提出基于图像块（patch-based）的自动目标分割方法，直接采用包含目标局部形态、外观信息的图像块集合来表示目标（这些图像块由预先分割好的图像训练得到），该方法通过图像块与图像中待分割目标的匹配以及分割结果所应具有的全局一致性约束相结合，实现了对马这类图像目标的分割。随后，Borenstein 又提出将基于图像块的高层信息与图像低层信

息结合来改进当仅利用高层信息或仅利用低层信息时分割结果不精确的问题。

Levin 等人[64]提出利用 CRF 对基于图像块的高层信息与低层信息同时进行训练，使得仅需少量结合低层信息的图像块就能达到以往仅考虑高层信息时数百块图像块才能得到精确分割。Shotton 等人[65]提出的 TextonBoost 是一种新的基于 texton（包括形状和纹理信息）的特征，并利用 Boosting 分类器对训练数据中目标的 texton 特征进行学习，最后通过将分类器结合到 CRF 中实现上下文相关的自动目标分割。Winn 等人[66]提出的自动目标分割方法基于局部组成部件的空间布局，考虑了相邻部件间的空间关系，允许部件的任意缩放。

图像目标分割方法与应用场景图像及应用目的有关，用于图像目标分割的场景信息也有亮度、色彩、纹理、结构、温度、频谱、运动、形状、位置、梯度和模型等。由于图像的多义性和复杂性，许多分割工作无法完全依靠计算机自动完成，而手工分割又存在工作量大、定位不准确的难题，因此，人们提出了一些人工操作和计算机自动定位相结合的方法，充分发挥各自优势，实现图像目标的快速分割。图 2-12 所示为刘陈博士设计的基于智能人机接口的即时过程式分割方法[57]，当用户驱动鼠标对目标对象的边界进行跟踪时，智能画笔将动态地根据图像局部统计特征估计每个即时时刻目标分割计算所需的待分割区域、外观样本、目标轮廓等即时局部信息，快速计算局部分割结果并及时反馈。



图 2-12　妇女图像即时交互分割过程截图及结果

图像分割算法的评估技术有很多种，一般通过建立统一的实验平台进行评估。给定一组测试数据以及其对应的真实分割结果（通常是人工确定的），在这组测试数据上进行实验并得到实际分割结果，然后通过比较计算耗时、错误率、整体一致性等指标，达到对分割算法的评估。考虑到评价的客观性，加州大学伯克利分校所发布的用于分割的标准图像库㊀，被许多科研工作者作为测试图像数据的主要来源。

㊀ http：//www. eecs. berkeley. edu/Research/Projects/ CS/vision/grouping/segbench。

2.3 目标的表示与描述

在图像预处理和图像分割之后，往往需要对得到的像素集用更为简单明确的数值、符号或图来表征，这些数值、符号或图是按一定的概念和公式从原目标区域中产生的，它们反映目标区域基本的重要信息和主要特性，这些数值、符号或图应有利于人或计算机对于图像目标的分析和理解，它们通常被称为目标的特征（整体特征）。

基本上，表示一个目标区域有两种选择：一是用其外部特性来表示区域（如它的边界）；二是用其内部特性来表示区域（如组成区域的像素）。当重点关注目标区域的形状特征（Shape Feature）时，可以选用外部表示法。而当重点关注区域内部性质时，可以选用内部表示法，比如光谱特征（Spectrum Feature）和纹理特征（Texture Feature）。有些情况下，同时使用上述两类表示方法。无论哪种情况，用整体特征进行目标的表示和描述，对尺寸变化、平移和旋转都不是很灵活的。

2.3.1 光谱特征

相对其他特征而言，光谱特征（也称颜色特征）具有描述简便直观的特点，而且对大小、方向都不敏感，在一些情况下表现出相当强的鲁棒性。人眼对彩色的分辨率高于对黑白图像的分辨率，因此彩色图像所携带的信息远远超过了灰度图像。但只用光谱特性很难完整而准确的描述一个具体物体，因为许多不同的目标所表现出的光谱特征可能相同（如全色遥感影像中机场与其他人文建筑等），这使得其应用容易受限。

1. 颜色空间

颜色空间又叫彩色空间、颜色模型，是用来表示颜色的三个参数所构成的3D空间，是颜色抽象表示和描述的方法，是在某些标准下用通常可接受的方式来简化的颜色规范。因此，颜色空间是进行颜色信息研究的理论基础。RGB、XYZ、HIS 和 $L^*a^*b^*$ 是四种不同的颜色空间，以不同的方式描述图像目标的颜色特征[67、68]。

人类能够感受到不同的颜色是由于视网膜中有三种不同的感受彩色的锥细胞，它们分别对应于红（R）、绿（G）、蓝（B）三种颜色。于是，人眼感知的所有颜色都可以看做是红、绿和蓝三原色的不同组合。RGB 颜色空间是最基本

的颜色空间，其他所有的颜色空间都是由它经过线性或非线性变换得出的。

XYZ 空间包含了所有人类能够感觉的颜色，这三种基色是虚拟的，使得颜色比配全部为正值，而且它是基于实验测定的颜色匹配函数，因此它不同于 RGB 颜色空间只是表示监视器所能显示的颜色范围，可以显示所有的颜色。

HIS（Hue Intensity Saturation）空间是从人的心理感知角度建立的，最能体现人眼的视觉特点。其中，H 是指一种颜色在色谱中所对应的主波长（色调），S 相当于颜色的纯度（饱和度），I 表示强度和亮度（密度）。

$L^*a^*b^*$ 颜色是从 RGB 模式转换为 HSB 模式和 CMYK 模式的桥梁。该颜色模式由一个发光率（Luminance）和两个颜色（a 和 b）轴组成，具有“独立于设备”的特性，即使用任何一种监视器或打印机，其颜色效果不变。

2. 颜色统计特性

利用光谱特征来识别目标，主要工具是单波段图像的灰度直方图（见图 2-13）和多光谱图像的颜色直方图。直方图的横轴表示颜色的等级，纵轴表示具有该颜色等级的像素在整个图像区域中所占的比例。直方图是图像区域中灰度等级或颜色等级出现次数的统计比较结果，不能反映某一像素色彩值的位置信息。

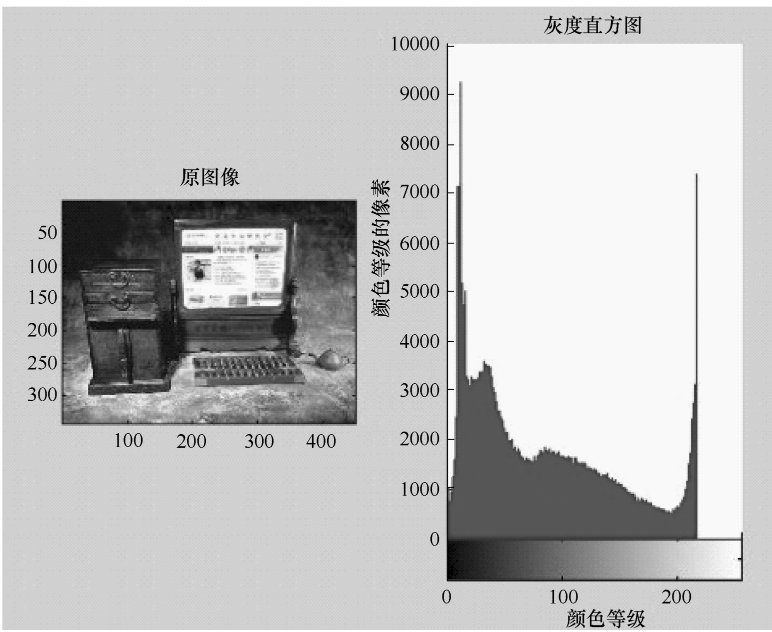


图 2-13　单波段图像的灰度直方图

除了颜色直方图之外，还有颜色矩（Color Moment）、颜色集（Color Sets）等其他一些颜色特征表示方式。颜色矩的数学基础是任何颜色分布均可由它的矩来刻画，并且由于大部分信息集中在低阶矩，色彩的统计低阶矩不仅能描述区域大众主要的色彩分量，而且可以反映出区域中的色彩分布情况。一阶矩对应色彩均值，二阶矩对应色彩标准差，三阶矩对应色彩偏度。

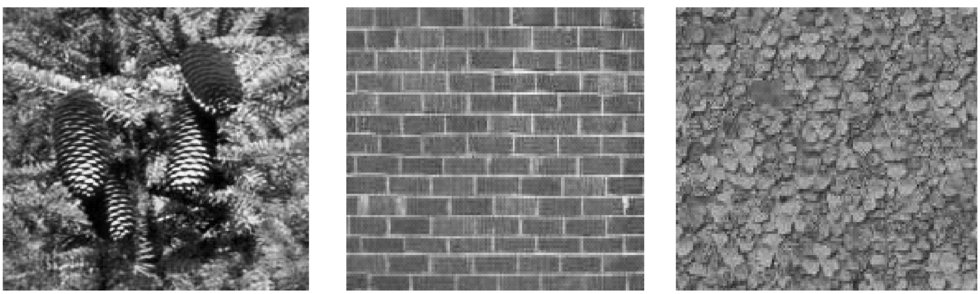
2.3.2　纹理特征

纹理是人类视觉系统对自然界物体表面现象的一种感知，是人们描述和区分不同物体的重要特征之一。如图 2-14 所示，常见的纹理有以下三种类型：

1）自然纹理。这类纹理来源于真实物体表面，大多呈现不规则性、随机性强。

2）人工合成纹理。是用计算机算法模拟或人为生成的表面纹理，一般形状规则、确定，分布均匀。

3）混合纹理。由人工纹理随机分布于物体表面或自然景物中构成[69]。



a)　　b)　　c)

图 2-14　常见纹理的示例

a）自然纹理　b）人工合成纹理　c）混合纹理

纹理最明显的视觉特征是粒度或粗糙性、方向性、重复性或周期性。同时，纹理是一个区域特征，与观察尺度相关。从人们的视觉感知来说，纹理有两个要素：引起视觉感知的像素灰度/颜色变化模式的基本单元，即纹理基元；纹理基元按一定规律排列，变现为某种规律性，也可以表现为随机性。所以，纹理特征可以认为是图像中灰度、颜色或细小的结构形状在空间上呈现规律的变化[70]。描述纹理的方法可以分为统计方法、结构化方法和基于模型的方法三大类。

1. 统计方法

从区域统计的角度去分析纹理图像的方法称之为基于统计的纹理分析方法，

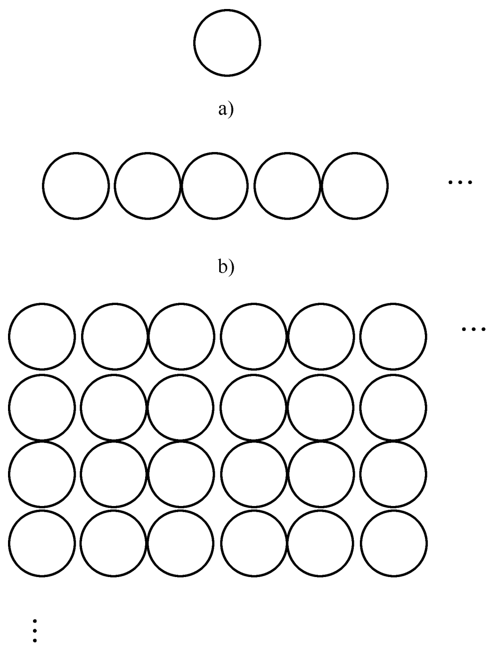
该类方法可以在空域和频域中进行。在图像空间域中包括矩、自相关函数、灰度共生矩阵、边缘频率、游程长度等；在频域中有频谱分析法。

基于空间域的纹理统计方法：纹理矩是与纹理基元形状和灰度空间分布有关的几何特征；空间自相关函数的基本思想是利用像素之间的灰度相似性计算描述图像纹理的规则度和粗糙度；灰度共生矩阵克服了直方图不反映空间位置信息的弱点，是图像灰度变化的二阶统计度量；边缘频率通过检测边缘分布的一阶和二阶统计量，可以度量出纹理的粗糙度、对比度、随机性、方向性等属性。

频谱技术利用傅里叶变换将空间域的纹理图像变换到频率域中，从而获得在空间域不易提取的纹理特征，主要用于通过识别频谱中高能量的窄波峰寻找图像中的整体周期性[8]。利用统计的方法对频率特性进行度量，可以派生出许多纹理特征的描述子（直方图、熵、均值、方差、斜度等）。

2. 结构化方法

结构化方法有两个步骤：一是纹理基元的提取；二是发现图像纹理中基元的排列规则。通常纹理基元由图像中具有均匀灰度的区域构成。纹理基元具有面积、周长、偏心率、方向、延伸度、矩等特征。结构化分析方法通常首先确定纹理基元，然后根据句法模式识别理论，利用形式语言对纹理的排列规则进行描述[68]。图 2-15 所示为纹理的结构化描述。



a)

b)

c)

图 2-15　纹理的结构化描述

a）纹理基元　b）由规则 $S\rightarrow aS$ 生成的模式

c）由 $S\rightarrow aS$ 和其他规则生成的二维纹理模式

结构化方法的优点是有利于对纹理构成的理解和高层检索使用，适合于描述人工规则纹理。而对于自然纹理来说，纹理分布的随机性使得纹理基元提取相当困难，基元之间的

排布规则不易用确定的数学模型描述。因此，结构化方法在随机纹理描述中应用不多。

3. 基于模型的方法

基于纹理模型的方法是通过所建立的图像模型来描述纹理的。常见的纹理模型方法有 Markov 随机场、自回归模型和分形维模型[69]。Markov 随机场（简称 MRF）是广泛使用的纹理模型，该模型在二维空间分析纹理图像的灰度变化，获得图像中局部空间上下文信息。自回归模型的系数表征纹理的特点和类型，对于粗纹理来说，自回归模型的邻域系数是相近的，而对于细纹理来说，自回归模型的邻域系数具有很大的不同。

许多自然物体表面在不同尺度上呈现粗糙性和自相似性，分形维模型是度量这些特性的有力工具。分形维模型的重要特征包括：分形维大小与人们对物体表面粗糙程度的视觉感知具有一致性，即光滑的物体表面具有较小的分形维值，而较为粗糙的表面具有较大的分形维值；分形维具有尺度不变性，物体表面的分形维模型广泛应用于物体的粗糙度、不规则性、自然纹理的分析。

2.3.3 形状特征

对于刚体目标来说，形状是其固有的一个本质特征，形状特征表达的一条重要准则是要求对目标的位移、旋转及尺度缩放具有不变性，因此利用形状特征来描述目标无疑是复杂背景下目标自动识别的一个重要方向。形状特征可以分为空间域几何特征和变换域几何特征两个大类。

1. 空间域几何特征

在经典的几何理论中，面积、周长、长度、宽度、主轴方向、凹凸面积、紧密度、实心度及偏心率、曲率这些特征得到了广泛应用。面积和周长可以很容易地从目标分割的过程中计算出来。面积是物体总尺寸的一个方便的度量，面积只与该物体的边界有关，而与其内部灰度级的变化无关。物体的周长在区别刚体目标时特别有用，一个形状简单的物体用相对较短的周长来包围它所占有的面积。

当一个目标从图像中分割出来后，计算它在水平和垂直方向的跨度也是很容易的，只需知道物体的最大和最小行/列号就可以了。但对具有随机走向的物体，水平和垂直并不一定是感兴趣的方向。在这种情况下，有必要确定物体的主轴并测量与之有关的长度和宽度。当物体的边界已知时，有几种方法可以确

定一个物体的主轴：可以算出物体内部点的一条最佳拟合直线（或曲线）；也可以从矩（Moments）的计算得出，关于矩的概念将在第 4 章 4.3 节讨论；应用物体的最小外接矩形（MER）也能进行计算[23]。

紧密度是在一定程度上描述区域紧凑型的全局性形状测度，当形状为圆时，紧密度为最小值 1，它是一个旋转、尺度及平移不变量，又是一个非矢量数值；偏心率为区域的主轴和次轴的比率，它区分不同宽度目标的能力比较强，长而窄的物体和短而宽的物体偏心率差别很大；曲率描述了边界上各点在边界方向上的变化情况，是人类视觉系统观察场景的重要线索，是从轮廓中提取出来的最为重要的特征值之一。

2. 变换域几何特征

因为不受待识别目标大小、位置、方位的影响，不变矩在图像目标识别方面得到了广泛的应用。最早是 Hu[71] 在 1962 年通过代数不变量引入矩不变量，再对几何矩进行非线性组合，进而得到一组对于图像平移、旋转、尺度不变的矩，并引入到模式识别领域。近年来经过许多学者的改进，使得不变矩特征的描述能力不断得到提高。

傅里叶描述子是经典的形状描述方法，易于实现，并且有坚实的数学理论基础。主要思想是：在提取目标之后，用角累加函数表示物体边界点集合，然后对角累加函数进行傅里叶变换，可以生成一个复系数集合，这些系数即为傅里叶描述子[72]。低频系数代表了一般的形状属性，高频系数则代表了形状细节，利用傅里叶系数还可以构造能直接反映区域形态的一些参数。

形状的小波表示方式在粗尺度上给出形状的全局信息，在细尺度上给出局部信息。由于小波变换提供了多分辨率表示，因此目标识别的技术方案可以根据输入图像的尺度灵活调整。如图 2-16 所示，对原图像（左上）进行了三个级别（尺度）的二维快速小波变换，结果是将其划分为子图像的集合[72]：在第一级小波变换时，原图像被划分为一个低频子图像 LL 和三个高频子图像（LH、HL、HH）；二级小波变换是对第一级得到的低频子图像 LL 进行递归分解的过程，第一级分解得到的三个高频子图像保持不变；更高级的小波分解以此类推。可见，低频子图像 LL 是原图像在低分辨率上的一个近似，剩余的三个子图像都包含高频成分，它们在不同的分辨率和方向上表示了原图像的高频细节。

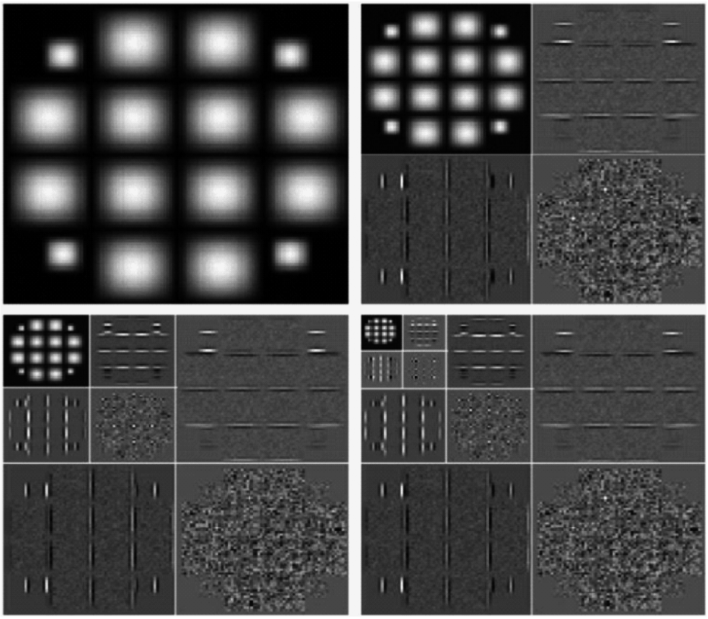


图 2-16　三个尺度下的二维小波变换

2.4　特征空间的优化

确定合适的特征空间是目标识别的一个关键的问题。如果所选用的特征空间能使同类目标分布具有紧致性，不同类别目标彼此分开，即各类样品能分布在该特征空间中彼此分隔开的区域内，这就为分类器设计提供良好的基础。反之，如果不同类别的样品在该特征空间中混杂在一起，再好的设计方法也无法提高分类器的准确性。对特征空间进行优化有两种基本方法：一种是特征选择，另一种就是特征变换。

2.4.1　特征选择

特征选择指对原始数据的特征进行筛选，保留那些对区分不同类别的必要特征，舍去那些对分类并无多大贡献的特征，使得最终的特征空间能够反映分类的本质。特征选择的方法按照特征选择过程与分类器之间的交互程度可以分为过滤式（Filter）、Wrapper[73]、嵌入式、混合式几种类型。

过滤式特征选择是完全独立于分类器的，这也是最常见的一种特征选择方式，选择过程计算量小，但是选择的特征不一定很适合分类。在 Wrapper 方法

中，特征子集的性能使用一个分类器在验证样本上的正确率来衡量，这样选择的特征比较适合该分类器，但不一定适合其他的分类器。由于在特征选择过程中要评价很多特征子集（子集的数量呈指数级增长），即使采用顺序前向搜索，Wrapper 方法的计算量都是很大的，只适合特征维数不太高的情况。Wrapper 方法的另一个问题是当训练样本较少时会造成过拟合，泛化性能变差。

嵌入式方法是在分类器的训练过程中包含了特征选择功能，因此跟 Wrapper 方法一样也是依赖于分类器的。一个经典的方法是 LASSO[74]，近来有代表性的两种嵌入式方法是稀疏支持向量机[75]和 Boosting 特征选择[76]。混合式特征选择结合不同的方法以实现更好的计算复杂性-分类性能的折中，在初始特征数量非常大时经常使用，很多此类方法[77]在三个阶段先后用三种方法削减特征个数：过滤、聚类、组合式选择。过滤方法和 Wrapper 方法也经常结合使用。

特征选择领域大部分的研究工作都集中在过滤式方法。模式识别领域早期的工作多把关注点放在搜索策略上，特征子集评价准则多采用基于高斯密度假设的距离准则，如 Fisher 准则、Mahalanobis 距离等。其实，特征子集的评价准则更为重要，当准则较好地衡量特征子集的可分性且比较稳定时，简单的搜索策略就能产生良好的分类性能。

特征选择常常面临着保留哪些描述量删除哪些描述量的抉择，信息论在这方面为图像识别提供了许多有用的方法[78-80]，如图像频率（Image Frequency，IF）、χ^2统计量（CHI）、术语强度（Term Strength，TS）、信息增益（Information Gain，IG）法和互信息（Mutual Information，MI）方法等。

基于图像频率的特征选择方法简单易行，可以在降低特征空间复杂度的同时去掉一部分噪声特征，但低频特征也可能带有很大的信息量，该方法直接去除低频特征会影响识别效果；χ^2 统计量度量特征和类别独立性的缺乏程度，优点是降维效果比较好，缺点则是统计花费大；术语强度的特点是基于目标聚类的方法，认为在相关目标中出现次数越多的特征具有信息量，这样可以去掉大部分无信息量或带有很少信息量的特征。

信息增益法[81]是依据某个特征项为整个分类所能提供的信息量多少来衡量该特征项的重要程度，从而决定对该特征项的取舍。理论上讲，信息增益应该是最好的特征选择方法，但实际上由于许多信息增益比较高的特征出现频率往往较低，所以当使用信息增益选择的特征数目比较少时，往往会存在数据稀疏

问题，此时识别效果也比较差。

互信息法的基本原则是选择类别相关的特征，同时排除冗余的特征。特征与类别之间的互信息很好地度量了特征的相关性，特征与特征之间的互信息则度量它们之间的相似性（冗余性）。因此，基于互信息的特征选择一般遵循这样一种模式：在顺序前向搜索中寻找与类别互信息最大而与前面已选特征互信息最小的特征项[82]。文献［83］提出的条件互信息用来度量在一个已选特征的条件下另一个新的候选特征对分类的相关性。文献［84］通过分析一种相关度设计一种快速的两步特征选择方法。虽然 Yang 等人[85]从数学的角度比较了信息增益法和互信息法，解释了实验结果的一些现象，但是，评价特征选择方法的标准并没有从理论上得到验证。

2.4.2　特征变换

特征变换是通过一种映射变换改造原特征空间，也就是说新的每一个特征是原有特征的一个函数。传统的线性变换方法主要有主分量分析（Principal Component Analysis，PCA）[86,87]、独立分量分析（Independent Component Analysis，ICA）[88]、线性判别分析（Linear Discriminant Analysis，LDA）[89,90]。

主分量分析的目的是寻找在最小均方意义下最能代表原始数据的投影方法，它通过 KL 变换得到互不相关的新特征分量，而且可以根据需要选取最主要的那部分，从而在降维的同时最大程度地保留了原始数据的信息；由于主分量分析假定数据集满足高斯分布，在非高斯分布的情况下常采用独立分量分析，而统计独立是比主分量分析所要求的不相关条件更加严格的条件，只有对于高斯随机变量，这两个条件才相同[31]；相对前两种方法寻找的是用来有效表示的主轴方向，线性判别分析方法寻找的是用来有效分类的方向[52]，该方法又叫 Fisher 判别分析，也是假设所有样本在总体上服从高斯分布，其目的是使子空间中类间离散度（Sb）和类内离散度（Sw）的行列式之比达到最大。另外，LDA 提取的特征个数受到类别数的限制，而当训练样本数相对特征维数较小时，Sw 为奇异，会带来很多计算上的问题。

由于非高斯分布、小样本问题的存在，特征变换也成为了近年来特征提取技术的一个热点，这方面工作主要可以分为以下几个方向：

1）针对小样本的线性特征提取方法；

2）类内协方差矩阵不同的情况下的异方差（heteroscedastic）判别分析；

3）非高斯分布下的特征变换方法；

4）局部空间特性保持的特征变换方法；

5）非线性特征变换方法；

6）二维模式特征变换方法。

小样本学习的一个典型例子是图像分类，如果直接用图像中所有像素点的值作为特征量，矢量的维数非常高，而每一类的样本数又很少。克服 Sw 奇异性的一个直接方法是正则化（regularized）判别分析[89]，通过矩阵平滑使 Sw 变得非奇异。Fisherface 方法则用 PCA 把特征维数从 D 降到 $N-M$（N 是样本数，M 是类别数）使 Sw 变得非奇异。但是，Sw 的维数由 D 降到 $N-M$ 会损失一些鉴别信息，而降到 $N-1$ 维则不会有损失。而这时 Sw 仍然是奇异的，就需要从 Sw 的零空间（对应本征值为 0）提取一些特征。与一般的 LDA 方法先对 Sw 对角化然后对 Sb 对角化相反，一种 Direct LDA 方法[91]先对 Sb 对角化后从变换后的 Sw 提取对应较小本征值的鉴别矢量。

对于类别协方差矩阵不同的情况异方差判别分析[92]方法可以得到比 LDA 更好的分类性能。对于非高斯分布或任意分布的情况，非参数判别分析是提取判别特征的一个基本思路，由此发展起来的方法还包括基于决策边界的判别分析。在不假设参数概率密度的情况下，也可以用分类性能准则直接对鉴别投影矢量进行优化，这样的准则如最小分类错误（MCE）和特征与类别之间的互信息[93]。对于每类样本为多模态分布的情况可以采用基于混合高斯密度的鉴别分析[94]。

局部空间特性不变的特征变换方法借鉴了流形学习（Manifold Learning）的思想，目的是在子空间中保持样本点之间的相邻关系。流形学习的问题是只对训练样本进行投影，要推广到测试样本就需要用一个参数模型或回归网络来表示投影的过程。He 等人[95]提出的局部性保持投影（LPP）方法通过优化一个局部性保持准则来估计投影矢量，可转换为矩阵本征值分解问题，LPP 是一种非监督学习方法，被推广到监督学习和核空间；Yan 等人[96]提出一种基于样本邻近关系分析的特征提取的统一框架，称为嵌入图（Embedded Graph），并在此基础上提出一种新的判别分析方法；另外，Isomap 流形学习方法[97]也被推广到监督学习用于非线性特征变换。

几乎所有的线性特征投影方法都可以推广到核空间。Schölkopf 等人[98]最先将核函数引入 PCA，提出 Kernel PCA（KPCA）方法；类似地，将核函数引入 Fisher 鉴别分析，提出了 Kernel FDA（KFDA）方法[99]；许多研究人员[100]在核空间中对 PCA 降维和 FDA 特征变换等技术进行深入分析，进而提出了许多算

法，比如 Kernel LPP 和 Kernel Direct LDA 等。

二维模式主成分分析（2D-PCA）[101]或判别分析（2D-LDA）[102]是近年提出的一种针对图像模式的特征变换方法。这类方法直接在图像矩阵上计算协方差（离散度）矩阵，该矩阵的维数等于图像的行数或列数，计算起来简便多了。另外，矩阵投影到每个本征矢量得到一个矢量，而不是一个值，这样得到的特征值个数也远远多于 LDA。在高维图像人脸识别实验中，2D-PCA 和 2D-LDA 的分类性能分别优于 PCA 和 LDA。二维变换方法实际上是基于图像行或列的变换方法，即对每一行或每一列分别投影得到特征，可以推广到基于图像块的投影。

在设计特征空间的初期阶段，应该尽量多地列举出各种可能与分类有关的特征，这样可以充分利用有用的信息。对此，Kanal 曾经总结过经验：样品数 N 与特征数 n 之比应足够大，通常样品数 N 是特征数 n 的 5 ~ 10 倍[103]。但高维度特征向量对后面的分类器存在不利的影响，很容易出现模式识别中的“维数灾难”现象。而且，并不是所有的特征项对分类都是有利的，很多提取出来的特征可能是噪声。因此，如何降低特征向量的维数，并尽量减少噪声，仍然是特征空间优化的两个关键问题。

特征选择和特征变换都是为了达到维数削减的目的，在降低分类器复杂度的同时可以提高分类的泛化性能。两者也经常结合起来使用，如先选择一个特征子集，然后对该子集进行变换。近年来为了适应越来越复杂（特征维数成千上万，概率密度偏离高斯分布）的分类问题的要求，不断提出新的特征空间优化方法，形成了新的研究热点。

2.5　本章小结

本章首先从认知科学的角度上介绍了图像目标特征提取的重要性及其在目标识别系统中的意义；然后给出了整体特征和局部特征的定义，并进行了简单的区分；接着着力阐述了图像目标分割的分割方法和研究现状，以及图像分割和目标分割的相互关系；最后对目标的各种表示与描述方法进行了详细的论述和比较，并综述了近些年来国内外学者在目标特征空间优化方面的科研进展。

特征提取和表示已经成为图像目标识别甚至机器视觉领域中的关键步骤，好的提取和表达方式能够极大地简化以及优化后续的处理过程。不过，有效的

目标分割和信息表示技术往往也需要对图像内容的认知学习和分析推理，这就衍生了交互式分割和交替式目标识别的思想，也是一个新的发展方向。总之，目前还没有一种完全自动的特征提取技术能适用于任何具体问题，一般需要在特定的任务中，甚至特定的图像里，选用一种或几种不同的方法，从而提取出合适的目标特征，完成图像目标识别以及场景理解等任务。

第 3 章　基于整体特征的目标识别

你们在想要攀登到科学顶峰之前，务必把科学的初步知识研究透彻。还没有充分领会前面的东西时，就绝不要动手搞往后的事情。

——巴甫洛夫·伊凡·彼德罗维奇（1849—1936）

3.1　引言

生物每天都在进行各种情况下的模式识别——如寻找食物、迁移、逃避敌害、辨认同伴等，这是生物与生俱来的应付周围环境所必需的能力，也是一种智能最常见的体现。当然，它可能只是很简单的本能，如微生物来到 pH 值不合适的环境中就会逃走；也可能需要训练和推理，如医生通过望闻问切或者借助仪器判断病症。

模式识别研究的目的是构造自动处理某些信息的机器系统，以代替人完成分类和辨识的任务。它的研究对象基本上可以概括为两类：一类是有直觉形象的如图片、相片、图案、文字等，一类是无直觉形象而只有数据或信号波形如语言、声音、心电脉冲、地震波等。但对模式识别来说，无论是数据、信号还是平面图形和物体，都是除掉它们的物理内容找出它们的共性，把具有同一共

性的归为一类，有另一种共性者则归为另一类[104]。

模式即描述子的组合，例如在第2章2.3节中讨论过的那些符号，在许多有关模式识别的著作中，也经常用特征来表示一个描述子。模式类是一个拥有某些共同性质的模式族。模式类用 ω_1，ω_2，…，ω_W 表示，这里 W 是模式类的数量。由机器完成的模式识别是对不同的模式分配各自所属类的技术，这种技术是自动的并且尽可能地减少人的介入[8]。

正如第1章1.3.2节所述，我们可以认为图像识别是图像处理与模式识别两个学科的结合，也可以把图像识别看做专门针对图像数据的模式识别。在本章中，我们称单个图像区域为目标、对象或者模式。

3.2　模式识别方法概述

模式识别方法具有多样性，对于如何将它们进行分类没有明确的定义。我们可以大体将其分为两个主要类型：决策理论和结构判别。实践中的三种常用模式组合——向量（用于定量描述）、串和树（用于结构描述），就是分别适用于这两类模式识别方法的。

1. 决策理论

这是一种数学方法，它是受数学中的决策理论的启发而产生的识别方法。它主要是建立在被研究对象的统计知识上，也就是对图像目标进行大量的统计分析，抽出图像中本质的特征而进行识别。在这种方法中很大的精力都集中在提取图像特征方面，也就是把图像目标大量的原始信息缩减为少数特征，然后再进行特征空间优化，将最终的模式向量作为识别的依据。

模式向量一般用黑体字母表示，比如 $\boldsymbol{X}$，$\boldsymbol{Y}$ 和 $\boldsymbol{Z}$，并采取下列形式：

$$\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}. \tag{3-1}$$

这里，每个分量 x_i 代表第 i 个描述子，n 是与模式有关的符号总数。模式向量是用列向量表示的，即 $n \times 1$ 阶矩阵。模式向量 $\boldsymbol{X}$ 中的元素性质取决于描述物体模式自身所采用的方法。

2. 结构判别

结构性方法，也称语言学方法。它是立足于分析图像结构，把一幅图像看成语言构造。例如一个英文句子，是词和短语组成的并按一定的语法表达出来，其中最基本元素是单词。与此类似，图像是由一些直线、斜线、点、弯曲线及环等组成。剖析这些基本元素，看它们是以什么规则构成图像，这就是结构分

析的课题。这些基本元素相当于句子中的单词，那些直线、曲线的组合相当于短语，它们全体如何构成图像就相当于语法规则。此时，图像识别就相当于检查图像所代表的某一类句型是否符合事先规定的语法，如果语法正确就识别出结果。

在某些应用中，模式的特性很适于用结构关系进行描述。例如，指纹识别基于称为细节的指纹特征的相互关系。综合指纹的相对大小和位置，这些特征是描述指纹纹路属性的主要分量，如指纹的端点、分支、合并以及不连续段。这类识别问题通常用结构性方法会得到很好的解决，因为它们的特征不仅与数量有关，而且各个特征间的空间关系也决定着它们的类别归属。

串的描述适于生成目标模式和其他实体模式，它们的结构是基于原始元素的较为简单的连接，通常和边界形状有关系。对许多应用来说，更有效的一种方法是树形描述结构，也就是一种主要的分层有序的结构。如图 3-1 所示，一张关于乡村风景的照片，树的根节点代表整幅图像，下一级节点表示此图由前景和背景构成，前景又由地面和非地面区域构成，再下一层进一步描述地面和非地面区域……可以一直继续这样的细分，直到到达在图像解析不同区域的能力极限。

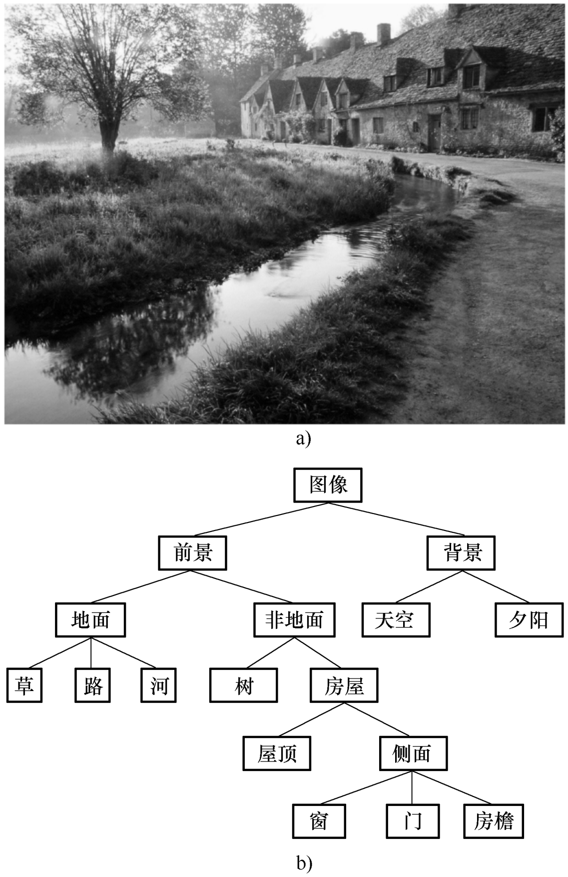


图 3-1　一幅图像的多级树形结构描述

a）乡村风景图像　b）树形描述结构

从上述两类方法看来，第 1 种方法（决策理论）没有利用图像本身的结构关系，第 2 种方法（结构判别）没有考虑图像目标受到的噪声干扰。如果两者

结合起来考虑可能会有新的识别方法，目前这方面的研究还不多。由于本书后续章节中的技术方法主要是针对复杂背景下的图像目标表示与识别，所以本章对结构性方法不做详细介绍，如果有读者比较感兴趣的话，可以查阅参考文献［8，54，67，104］中的相关内容。

其实，模式识别也可以分为模式匹配和模式分类两个方向，与此相应的图像目标识别系统都由两个过程组成，即设计与实现。设计是指用一定数量的样本（叫做训练集或学习集）进行分类器或模型库的设计；实现是指用所设计的分类器或模型库对待识别的样本进行分类决策[54]。

目标分类一般需要构造有效的特征向量和充分利用相关领域的知识，而在许多实际应用中，很难得到有关特征概率和类别概率的先验知识，或者得到的数据不足以设计分类器。在这种情况下，可以使用模型直接匹配未知物体，并选择最佳匹配为最终分类结果[105]。

3.3 目标匹配的研究现状

利用特征进行模式匹配是目前目标匹配识别中最常用也最有效的方法，其具体含义是指图像中目标的特征同模型库中的模型相匹配。在许多图像目标识别任务中，待识别的目标数量较多，每一个目标拥有的特征也有许多，因此，在建立识别系统的时候，必须考虑特征的有效性和匹配算法的高效率。

3.3.1 两种目标匹配方式

可以这样定义目标匹配识别：给定一幅包含一个或多个物体的图像和一组对应物体模型的标记，系统应将标记正确地分配给图像中对应的物体或区域集合。对应于向量（定量描述）与串和树（结构描述）的模式组合形式，一般采用直接匹配和符号匹配两类方式。

1. 直接匹配

假设每一个特征类别是由它的特征来表示的。即假设第 i 类物体的第 j 个特征值表示为 f_{ij}。对于一个未知物体，其特征表示为 u_j。该物体和第 i 类的相似性由下式给出：

$$S_i = \sum_{j=1} w_j s_j \tag{3-2}$$

式中，w_j 是第 j 个特征的权值，权值的选择是以特征的相对重要性为基础的；第 j 个特征相似值是 s_j，它可以是绝对差、规范化差或其他距离测量值。最常用的

方法是用下式并考虑同特征一起使用的权值规范化：

$$s_j = |u_j - f_{ij}| \tag{3-3}$$

如果 S_k 是最高相似度值，则标记物体为 k 类。在此方法中，使用的特征可能是局部的，也可能是全局的。注意此方法没有使用特征之间的任何联系。

2. 符号匹配

一个物体不仅可以用它的特征来表示，而且可以用特征之间的联系来表示。特征之间的关系可以是空间的，或者是其他形式的。在这样的情况下，物体可能被表示为一个图形。图形的每一个节点都表示一个物体，弧线连接节点表示物体之间的联系。因此，物体识别问题可以认为是图形匹配问题。

一个图形匹配问题可以定义如下：有两个图形 G_1 和 G_2，包含 N_{ij} 个节点，其中 i 表示图形数，j 表示节点数，节点 j 和节点 k 之间的联系表示为 R_{jk}。在图形上定义一个相似性测量值，该测量值包含了所有节点和函数的相似性。

在目标识别的多数应用中，待识别的物体可能是部分可见的。因此，一个识别系统必须能从物体的部分视图来识别它们。那些使用整体特征和要求所有特征都存在的识别方法在这些应用中是行不通的。在某种意义上，部分视图识别问题和图形学中研究的图形嵌入问题是类似的。但当我们开始考虑节点相似性和节点之间关系时，物体识别中的问题与图形学问题就不同了。

3.3.2　匹配的相似度度量

对目标进行匹配识别，需要选用合适的相似度比较函数，这个函数可以称之为相似度度量。相似度度量具有特征依赖性，不同的特征应该采用不同的度量方法获得最佳的测度效果。由于局部特征是采用模式向量的方式描述的，计算两个特征向量之间的距离是它们相似度的一种很好的度量。设 d 为距离函数，$\boldsymbol{X}$，$\boldsymbol{Y}$，$\boldsymbol{Z}$ 为局部特征的模式向量，表示形式为 $\boldsymbol{X}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$。通常情况下，距离度量函数应该满足如下四个性质：

自相似性　$$d(\boldsymbol{X}, \boldsymbol{X}) = d(\boldsymbol{Y}, \boldsymbol{Y}) = d(\boldsymbol{Z}, \boldsymbol{Z}) = 0 \tag{3-4}$$

最小性　$$d(\boldsymbol{X}, \boldsymbol{Y}) \geqslant d(\boldsymbol{X}, \boldsymbol{X}) \geqslant 0 \tag{3-5}$$

对称性　$$d(\boldsymbol{X}, \boldsymbol{Y}) = d(\boldsymbol{Y}, \boldsymbol{X}) \tag{3-6}$$

三角不等性　$$d(\boldsymbol{X}, \boldsymbol{Y}) + d(\boldsymbol{Y}, \boldsymbol{Z}) \geqslant d(\boldsymbol{X}, \boldsymbol{Z}) \tag{3-7}$$

在实际应用中，所采用的相似度比较函数并不一定全都要满足上述的四条定理，可能只满足其中的一个或者几个。目前常用的距离函数有明可夫斯基距离、马氏距离、二次型距离和 EMD 距离等。

1. 明可夫斯基距离（Minkowski Distance）

$$D(\boldsymbol{X},\boldsymbol{Y})=L_p(\boldsymbol{X},\ \boldsymbol{Y})=\left[\sum_{i=1}^{n}|x_i-y_i|^p\right]^{\frac{1}{p}},\ p\geqslant 1 \tag{3-8}$$

当 $p=1$ 时，$L_1(\boldsymbol{X},\ \boldsymbol{Y})$ 称为海明距离（Haming Distance）：

$$L_1(\boldsymbol{X},\ \boldsymbol{Y})=\sum_{i=1}^{n}|x_i-y_i| \tag{3-9}$$

当 $p=2$ 时，$L_2(\boldsymbol{X},\ \boldsymbol{Y})$ 称为欧氏距离（Euclidean Distance）：

$$L_2(\boldsymbol{X},\ \boldsymbol{Y})=\left[\sum_{i=1}^{n}(x_i-y_i)^2\right]^{\frac{1}{2}} \tag{3-10}$$

当 $p\to\infty$ 时，$L_\infty(\boldsymbol{X},\ \boldsymbol{Y})$ 称为切比雪夫距离（Chebychv Distance）：

$$L_\infty(\boldsymbol{X},\ \boldsymbol{Y})=\max_{1\leqslant i\leqslant n}|x_i-y_i| \tag{3-11}$$

从向量范数的角度来讲，明可夫斯基距离可以称之为 p – 范数，海明距离、欧氏距离和切比雪夫距离分别称为 1 – 范数、2 – 范数和∞ – 范数[106]。

2. 马氏距离（Mahalanobis Distance）

马氏距离，即马哈拉诺比斯距离，是由印度统计学家马哈拉诺比斯提出的，表示数据的协方差距离。它是一种有效的计算两个未知样本集的相似度的方法，与欧氏距离不同的是它考虑到各种特性之间的联系（例如：一条关于身高的信息会带来一条关于体重的信息，因为两者是有关联的）并且是尺度无关的（Scale-invariant），即独立于测量尺度。其数学表达式为

$$D(\boldsymbol{X},\ \boldsymbol{Y})=\sqrt{(\boldsymbol{X}-\boldsymbol{Y})^{\mathrm{T}}\mathrm{C}^{-1}(\boldsymbol{X}-\boldsymbol{Y})} \tag{3-12}$$

其中，C 为特征向量的协方差矩阵，T 表示矩阵的转置运算。如果协方差矩阵为单位矩阵，马氏距离就被简化为欧氏距离；如果协方差矩阵为对角阵，则其也可称为正规化的欧氏距离。

马氏距离有很多优点。它不受量纲的影响，两点之间的马氏距离与原始数据的测量单位无关；由标准化数据和中心化数据（即原始数据与均值之差）计算出的两点之间的马氏距离相同。马氏距离还可以排除变量之间的相关性的干扰。它的缺点是夸大了变化微小的变量的作用。

3. 二次型距离（Quadratic Distance）

明可夫斯基距离对所有的特征向量平均对待，没有考虑特征向量之间的关系。二次型距离与马氏距离一样，考虑了各个特征向量之间的关联性。其数学表达式为

$$D(\boldsymbol{X},\ \boldsymbol{Y})=\sqrt{(\boldsymbol{X}-\boldsymbol{Y})^{\mathrm{T}}A(\boldsymbol{X}-\boldsymbol{Y})} \tag{3-13}$$

其中 $A = [a_{ij}]$ 为一个对称矩阵，表示特征向量之间的相关性；a_{ij} 为下标为 i 和 j 的特征分量之间的相似性。二次型距离考虑到特征分量之间的相关性，但是对称矩阵的计算量较大。

4. EMD（Earth Mover's Distance）

EMD 度量是 Rubner 等人[107]提出的一种相似度度量，它把运筹学的运输问题引入到图像识别中，采用最优化求解最小运输成本的方法来度量图像间的相似性。

在理解 EMD 计算原理时，可以把多个分布的其中之一视为地球表面的高山，另一分布则视为地球表面的低洼部分，而 EMD 主要的目的是要找出可以将低洼部分填平的最小成本。对地距离（Ground Distance）是用于计算高山与低洼部分的距离，也就是搬移一个单位所需花费的成本，当 EMD 的值愈小时则表示这个分布愈相似。计算 EMD 距离的方法比较复杂，不同应用需根据要求选择有效的对地距离[108]。

EMD 距离在最近得到了较广泛的关注，因为它能以一种非常自然的方式处理部分匹配的问题，对于处理图像领域中广泛存在的遮挡、轮廓片段匹配具有很大的用途；另外，当对地距离具有感知意义时，EMD 距离往往最能体现视觉感知上的相似性。

3.4　目标分类的研究现状

目标分类也可以称为模式分类，就是在特征空间中用统计方法把被识别对象归为某一类别。基本做法是在样本训练集基础上确定某个判决规则，使按这种判决规则对被识别对象进行分类所造成的错误识别率最小或引起的损失最小[54]。

模式分类不同于经典的统计“假设检验”技术，后者根据输入数据，判断零假设（或原假设、空假设）H_0与备择假设 H_1 中哪一个成立。简单地说，如果在零假设 H_0成立的前提下获得相应实际输入数据的概率小于某个“显著性水平”，则我们拒绝零假设 H_0而接受备择假设 H_1。模式分类也不同于严格意义上的“图像处理”。在图像处理中，输入的是一幅图像，输出的也是图像。图像处理的步骤常包括图像旋转、对比度增强和其他能保持所有原始信息的图像变换。而特征提取，比如检出图像中的峰谷点，将要损失信息。

如上所述，特征提取器输入模式，而输出特征值。特征的数目几乎总是少

于用于描述完整的感兴趣的目标所需的数据量，因而在这个过程中产生信息损失。而“联想存储器”的功能是输入模式，激发出另外一类模式。这个过程也损失信息，但损失的分量远比不上模式分类器所为。简而言之，因为决策在模式判别信息中至关重要的作用，所以它本质上就是一个信息压缩过程，不可能仅仅根据已知某个模式的类别隶属就重构该特定模式。分类过程中，信息量的损失更大，将原来图像中成千上万比特的像素颜色信息压缩至几个比特表示的类别信息。

另外还有 3 种密切相关的技术——回归分析，函数内插，和（概率）密度统计[52]，也经常要用到模式识别系统中的第一个步骤，不管是显式的运用或隐含的运用。回归分析的目的是对输入数据找到合适的函数表示，常用于预测新数据的值，其中线性回归的函数形式对输入数据而言是线性的，是到目前为止最流行也是研究最透彻的一种回归形式；在函数内插中，我们已知的（或者容易得出的）是一定范围内的输入数据对应的函数值，而要解决的问题是如何求出位于这些输入点之间的数据点的函数值；密度函数估计用于求解具有某种特定特征的类别成员（样本）出现的（概率）密度问题。

3.4.1 分类器设计技术

设计分类器是目标分类的主要任务和核心研究内容之一。分类器设计就是在训练样本集合上进行优化（如使每一类样本的表达误差最小或使不同类别样本的分类误差最小）的过程，也就是一个机器学习过程。下面将从不同的角度对图像目标识别常用的分类器进行归类，进而介绍它们的研究现状。

1. 按照分类器的数目

按照分类器的数目，可以分为单分类器方法和多分类器方法。顾名思义，单分类器方法中，全部目标类别共用一个分类器，多分类器方法为每个类别设置一个分类器。但是多分类器方法会带来一个很严重的“拒识”问题。如果某个目标和全部目标类别的相似度都小于相应的阈值，就无法识别该目标。这种情况下，还得调用单分类器方法，将其类别设置为相似度最大的那个类别。所以，为每个类别设置一个分类器的方法应用并不广泛。

还有一种思路，就是用多个弱分类器来联合投票进行目标识别，采用这种思路的多分类器方法被认为是结合不同分类器的优点、克服单个分类器性能不足的一个有效途径。其核心思想是，k 个专家判断的有效组合应该优于某个专家

个人的判断结果。投票算法主要有两种：Bagging 算法[109]和 Boosting 算法[110]，它们都是通过对训练样本集进行重采样或加权来训练多分类器的。不过，Bagging 算法是并行的，而 Boosting 算法是串行的，它们在训练每个分量分类器时，训练样本的抽取方式也有所不同。Boosting 方法作为一种集成机器学习方法，通过粗糙的、不太正确的、简单的、单凭经验的初级预测方法（弱分类器），按照一定的规则（在自组织自学习的方式下设计各弱分类器的权重），最终得出一个复杂的、精确度很高的预测方法（提升分类模型来解决复杂问题）。基于 Boosting 方法有许多不同的变形，其中 AdaBoost 方法[111]由于算法简单、运算速度快而被广泛应用于字符识别和人脸检测等领域。

与其他学习方法对样本集或特征集进行分解不同的是，纠错输出编码（Error-correcting output codes，ECOC）[112]是对类别集进行分解，通过组合多个二类分类器（这里的一类可以是一个类别子集）来实现多类分类。另外一种通过二类分类器实现多类分类的方法是把一对样本之间的关系分为“同类”（Intra-class）和“不同类”（Extra-class）两类，输入特征从两个样本提取（如两个样本对应特征的差），二类分类器的输出给出两个样本“同类”的概率或相似度，多类问题采用近邻规则进行分类。这种方法可以克服训练样本不足的问题，而且在训练后可任意增加或减少类别而不必重新训练，近年来已广泛用于人脸识别等生物特征识别问题。

2. 按照分类器训练过程中的人工参与程度

按照分类器训练过程中的人工参与程度，一般可以分为有监督（Supervied）和无监督（Unsupervied）识别。它们从本质上的区别就在于训练数据是否有已知的类别标签。无监督识别主要用于确定两个特征向量之间的“相似度”以及合适的测度，并选择一个算法方案，基于选定的相似度测度对向量进行聚类（分组）。通常，不同的算法方案可能导致不同的结果，这一点必须由专家进行解释[31]；而有监督识别可以通过学习有标签的数据，挖掘已知信息来设计分类器，能够以较小的训练集获得较高精度的模型。

对于海量的图像数据进行人工标注，浪费资源且不切实际，近年来，将标注数据和未标注数据结合起来用于目标识别受到广泛的关注，这就是半监督（Semi-supervied）识别方法[46]。Cohen[47]，Yao[48]和 Li[49]等学者分别将半监督识别应用到了人脸识别、航拍图像的目标检测以及图像分类等领域，取得了一些成果。

3. 按照分类器的数学模型

按照分类器的数学模型，可以分为生成（Generative）方法和判别（Discriminative）方法。生成方法中的朴素贝叶斯（Naive Bayes）分类器[54]是根据目标属于不同类别的概率来进行分类的，它将分类器设计问题转化为概率密度估计问题，给出了最一般情况下适用的“最优”分类器设计方法，该方法对各种不同的分类器设计技术在理论上都有指导意义；在判别方法中将每个目标表示为特征向量，进而视作整个特征空间的一个点，认为不同的类别是特征空间中不同区域或子空间，因此如果能够找到一个分离函数把属于不同类别的点分来，则识别任务就完成了，这种方法不依赖于条件概率密度的知识，其中最具代表性的是神经网络（Neural Network，NNet）[113]和支持向量机（Support Vector Machine，SVM）[114]。

混合生成-判别学习的识别方法[115,116]近年来受到了广泛的关注。这种方法结合了生成模型和判别模型的优点，一般先是对每一类模式建立一个生成模型（概率密度模型或结构模型），然后用判别学习准则对生成模型的参数进行优化。学习的准则可以是生成模型学习准则（如最大似然准则）和判别学习准则（如条件似然度）的加权组合[117,118]。结合判别学习的贝叶斯网络[119,120]也可以看做是混合-判别学习模型。

Jain 等人[121]把分类器分为基于相似度（距离度量）的分类器、基于概率密度的分类器、基于决策边界的分类器。第一种分类器常用于目标匹配识别，其性能取决于相似度或距离度量的设计，后两种分类器基本对应于生成模型和判别模型。此外，强化学习近年来在模式识别领域得到了深入的研究和广泛的应用[122,123]。它在本质上是一种在线学习，与有监督学习的最明显区别是不需要指明目标类别的标签，只需要外界对这次分类任务完成情况给出“对”或“错”的反馈。

3.4.2 性能评估方法

性能评估是目标识别系统设计的一个重要部分，它将决定系统是否满足特定应用的要求以及预期的作用。如果没有达到要求，设计者应当根据评估结果重新考虑和设计系统。另外，在特征选择阶段，错误分类概率也可以作为性能指标来选择特定分类器的最佳特征。

假设一个目标识别系统输出的各种结果统计情况见表 3-1。

基于此表，可以得到系统的查准率（Precision）、查全率（Recall）、正确率

（Accuracy）、错误率（Error）和 F-测度值的计算公式，即

表 3-1　目标识别系统输出结果

系统对两者关系的判断 \ 目标与类别的实际关系	属于	不属于
标记为 YES	TP	FP
标记为 NO	FN	TN

查准率 $$\text{Precision} = \frac{TP}{TP + FP} \tag{3-14}$$

查全率 $$\text{Recall} = \frac{TP}{TP + FN} \tag{3-15}$$

正确率 $$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3-16}$$

错误率 $$\text{Error} = \frac{FP + FN}{TP + TN + FP + FN} \tag{3-17}$$

F-测度值 $$F_\beta = \frac{(\beta^2 + 1) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \tag{3-18}$$

上面公式中，正确率和错误率不是很常用，因为计算公式的分母太大，导致其对识别正确的目标（TP）数目变化不是很敏感。F-测度值中的 β 是调整查准率和查全率在评价函数中所占比重的参数，通常采用 $\beta = 1$ 的 F_1 测度值。

对于目标识别系统来说，查准率和查全率是一对相互矛盾的物理量。提高查准率往往要牺牲一定的查全率，反之亦然。为更全面地反映分类系统的性能，一种做法是选取查准率和查全率相等时的值来表示系统的性能，该值叫做平衡点（Break-even Point，BEP）值。在找不到查准率和查全率相等的时候，可以取最接近的查准率和查全率的平均值作为 BEP 值。

对于分类的总体性能评估，有宏平均（Maro-averaging）和微平均（Micro-averaging）两种评估方式。宏平均是先计算每个类别的指标，再计算每个类别指标的平均值；微平均计算所有个体样本指标的平均值。显然，宏平均把类别作为最小的评价单位；微平均把个体样本作为最小评价单位。当样本在所有类别中分布均匀时，宏平均等于微平均；当每个类别的个体样本数目悬殊时，宏平均会和微平均有较大的差别。

近年来，信号检测领域中的 ROC（Receiver Operating Characteristics）曲线被引入到对分类识别的效果评估和优化中[124,125]。曲线图的 Y 轴和 X 轴分别是评价指标 TPR（True Positive Rate）和 FPR（False Positive rate），其中，TPR 和 FPR

的计算公式如下：

$$\mathrm{TPR}=\frac{TP}{TP+FN},\qquad \mathrm{FPR}=\frac{FP}{FP+TN} \tag{3-19}$$

随着阈值参数的调整，ROC 空间中的曲线不但能直观反映识别系统的性能，曲线下的面积 AUC（Area Under Curve）更可以量化分类器接受正例的倾向性。另外，ROC 空间对样本在类别间的分布不敏感，可以反映错误代价（Error Cost）等指标的变化，具有特别的优势。基于该曲线图的相等错误率（Equal Error Rate，EER）即为 TPR = 1 − FPR。

在目标检测领域中，将背景噪声正确地排除在目标类别之外的数目（TN）相对于正确检测出目标区域的数目（TP）来说过于庞大，而且它的计算对于检测系统的评估意义不大。目标检测系统更加关注于是否将目标全部检测出来以及检测出的区域有多少是虚警[79]，这就引出了 RPC 曲线图（Recall Precision Curves)，其 Y 轴和 X 轴分别对应评价指标查全率和虚警率（1-Precision)。有效地将 RPC 曲线用于目标检测系统的评价、比较以及优化，成为近期的一个研究热点[126,127]。

3.5　典型的图像目标分类器

聚类、朴素贝叶斯分类器、神经网络、支持向量机等，均是图像目标分类器的典型，将在本节做以详细介绍，并在后面章节中进行应用。分类器模型和学习方法多种多样，性能各有特点，一般来说，SVM 和 Boosting 在大部分情况下分类性能优异，但也有他们自身的不足——SVM 的核函数选择和 Boosting 的弱分类器选择对性能影响很大，分类的计算复杂度较高（如 SVM 的支持向量个数往往很大)。

3.5.1　基于聚类分析的分类器

作为统计学的一个分支，聚类就是将数据对象分组成为多个类或簇（Cluster)，在同一个簇中的对象之间具有较高的相似度，而不同簇中的对象差别较大。相异度是根据描述对象的属性值来计算的，距离是最常采用的度量方式(见本书 3. 3. 2 节)。

如图 3-2 所示，在机器学习领域中，聚类是典型的无监督学习（Unsupervised Learning)，不依赖预先定义的类别和带类标号的训练实例，也可以称之为观察式学习。基于聚类分析的分类与后面几节所述的有监督学习分类的不同之

处在于，它要划分的类是未知的，也就是说事先并不知晓要把目标分为哪几个具体的类别。

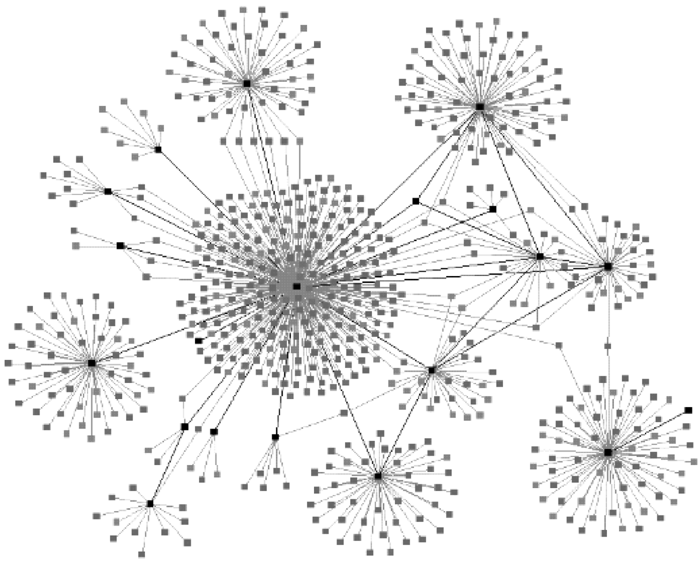


图3-2　聚类在最小世界网络（Small World Network）中的应用

聚类算法的选择取决于数据的类型、聚类的目的和应用，可以对同样的数据尝试多种算法，以发现数据可能揭示的结果。主要的聚类算法大体上可以划分为如下几类：

1. 划分的方法（Partitioning Method）

给定一个包含 n 个目标对象的数据集，一个划分方法构建对象数据的 k 个划分，每个划分表示一个类，并且 $k \leqslant n$。也就是说，它将样本划分为 k 个组，同时满足如下的要求：每个组至少包含一个对象；每个对象必须属于且只属于一个组。注意在某些模糊划分技术中第二个要求可以放宽。给定 k，即要构建的划分的数目，划分方法首先创建一个初始划分。然后采用一种迭代的重定位技术，尝试通过对象在划分间移动来改进划分。一个好的划分的一般准则是：在同一个类中的对象之间的距离尽可能小，而不同类中的对象之间的距离尽可能大。还有许多其他划分质量的评判准则。

为了达到全局最优，基于划分的聚类会要求穷举所有可能的划分。实际上，绝大多数应用采用了以下两个比较流行的启发式方法：k-平均值（k-means）算法，在该算法中，每个簇用该簇中对象的平均值来表示；k-中心点（k-medoids）算法，在该算法中，每个簇用接近聚类中心的一个对象来表示。这些启发式聚

类方法对在中小规模的数据集中发现球状簇很适用。为了对大规模的数据集进行聚类，以及处理复杂形状的聚类，基于划分的方法需要进一步的扩展。

2. 层次的方法（Hierarchical Method）

层次的方法对给定数据集进行层次的分解。根据层次的分解如何形成，层次的方法可以被分为凝聚的和分裂的方法。凝聚的方法，也称为自底向上的方法，一开始将每个对象作为单独的一个组，然后继续地合并相近的对象或组，直到所有的组合并为一个（层次的最上层），或者达到一个终止条件。分裂的方法，也称为自顶向下的方法，一开始将所有的对象置于一个簇中。在迭代的每一步中，一个簇被分裂为更小的簇，直到最终每个对象在单独的一个簇中，或者达到一个终止条件。

层次的方法的缺陷在于，一旦一个步骤（合并或分裂）完成，它就不能被撤销。这个严格规定是有用的，所示不用担心组合数目的不同选择，计算代价会较小。但是，该技术的一个主要问题是它不能更正错误的决定。有两种方法可以改进层次聚类的结果：一种是在每层划分中，仔细分析对象间的连接，例如 CURE 和 Chameleon 中的做法；另一种是综合层次凝聚和迭代的重定位方法，首先用自底向上的层次算法，然后用迭代的重定位来改进结果，例如在 BIRCH 中的方法[128]。

3. 基于密度的方法（Density-based Method）

绝大多数划分方法基于对象之间的距离进行聚类。这样的方法只能发现球状的簇，而在发现任意形状的簇上遇到了困难。随之提出了基于密度的另一类聚类方法，其主要思想是：只要临近区域的密度（对象或数据点的数目）超过某个阈值，就继续聚类。也就是说，对给定类中的每个数据点，在一个给定范围的区域中必须包含至少某个数目的点。这样的方法可以用来过滤“噪声”数据，发现任意形状的簇。DBSCAN 是一个有代表性的基于密度的方法，它根据一个密度阈值来控制簇的增长。OPTICS 是另一个基于密度的方法，它为自动的和交互的聚类分析计算一个聚类顺序[128]。

4. 基于网格的方法（Grid-based Method）

基于网格的方法把对象空间量化为有限数目的单元，形成了一个网格结构。所有的聚类操作都在这个网格结构（即量化的空间）上进行。这种方法的主要优点是它的处理速度很快，其处理时间独立于数据对象的数目，只与量化空间中每一维的单元数目有关。STING 是基于网格方法的一个典型例子，而 CLIQUE 和 WaveCluster 这两种算法既是基于网格的，又是基于密度的[128]。

5. 基于模型的方法（Model-based Method）

基于模型的方法为每个簇假定了一个模型，寻找数据对给定模型的最佳匹配。一个基于模型的算法可能通过构建反映数据点空间分布的密度函数来定位聚类。它也基于标准的统计数字自动决定聚类的数目，考虑“噪声”数据和孤立点，从而产生健壮的聚类方法。

COBWEB 是一个常用的且简单的增量式概念聚类方法，它的输入对象是采用符号量（属性-值）对来加以描述的，采用分类树的形式来创建一个层次聚类；CLASSIT 是 COBWEB 的另一个版本，可以对连续取值属性进行增量式聚类，它为每个节点中的每个属性保存相应的连续正态分布（均值与方差）；并利用一个改进的分类能力描述方法，即不像 COBWEB 那样计算离散属性（取值）和而是对连续属性求积分。

一些聚类算法集成了多种聚类方法的思想，所以有时将某个给定的算法划分为属于某类聚类方法是很困难的。此外，某些应用可能有特定的聚类标准，要求综合多个聚类技术。

传统的聚类方法已经比较成功地解决了低维数据的聚类问题，但在高维数据集中进行聚类时，却遇到了两个难以解决的问题：一是高维数据集中存在大量无关的属性使得在所有维中存在簇的可能性几乎为零；二是高维空间中数据较低维空间中数据分布要稀疏，其中数据间距离几乎相等是普遍现象，而传统聚类方法是基于距离进行聚类的，但在高维空间中无法基于距离来构建簇。

高维数据聚类分析是聚类分析中一个非常活跃的领域，同时也是一个具有挑战性的工作。信息技术的进步使得数据收集变得越来越容易，导致数据库规模越来越大、复杂性越来越高，如各种类型的贸易交易数据、Web 文档、基因表达数据等，它们的维度（属性）通常可以达到成千上万维，甚至更高。目前，高维数据聚类分析在市场分析、信息安全、金融、娱乐、反恐等方面都有很广泛的应用。在图像目标识别方面，随着图像内容越来越丰富以及特征描述子的维度不断增加，进行高维数据聚类分析已经提上日程。

3.5.2　基于朴素贝叶斯的分类器

朴素贝叶斯分类器进行目标分类的基本思想是利用特征项（特征分量）和类别的联合概率来估计给定目标的类别概率。该模型假定特征向量的各个分量间对于决策变量时相对独立的，即目标是基于特征项的一元模型，当前项的出现依赖于目标类别但不依赖于其他特征项。

训练集中的每个样本可以用一个 n 维特征向量 $\boldsymbol{V}=\{t_1, t_2, \cdots, t_n, C_i\}$表示，其中，$C_i$是类别标记，$1\leqslant i\leqslant m$，$t_k$ 是特征项，$1\leqslant k\leqslant n$。进行分类时，目标 T 被标记为 C_i，当且仅当

$$P(C_i|T)>P(C_j|T),1\leqslant j\leqslant m,i\neq j \tag{3-20}$$

根据概率理论的贝叶斯公式可知 $P(A|\ =B)=[P(A)P(B|A)]/P(B)$。应用此表达式，$P(C_i|T)$ 的计算可以表达为

$$P(C_i|T)=\frac{P(C_i)P(T|C_i)}{P(T)} \tag{3-21}$$

其中，$P(C_i)$ 为 C_i类目标的出现概率，其计算比较简单。在 n 分类中，如果训练集里各个类别的样本数目相同，则 $P(C_i)$ 可以取 $1/n$。$P(T|C_i)$和 $P(T)$ 的具体实现，通常又分为两种模型。

1. 多元伯努利模型（Multi-variate Bernouli Model）

目标 T 采用 DF 向量表示法[129]，即模式向量 $\boldsymbol{V}$ 的每个分量都是一个布尔值，0 表示相应的特征项在该目标中未出现，1 表示特征项在目标中出现。在这种方法中

$$P(T|C_i)=\prod_{t_k\in V}P(t_k|C_i) \tag{3-22}$$

$$P(T)=\sum_i\left[P(C_i)\prod_{t_k\in V}P(t_k|C_i)\right] \tag{3-23}$$

因此

$$P(C_i|T)=\frac{P(C_i)\prod_{t_k\in V}P(t_k|C_i)}{\sum_i\left[P(C_i)\prod_{t_k\in V}P(t_k|C_i)\right]} \tag{3-24}$$

其中，$P(t_k|C_i)$是对 C_i 类目标中特征 t_k出现的条件概率的拉普拉斯估计：

$$P(t_k|C_i)=\frac{1+N(t_k,C_i)}{M+N(C_i)} \tag{3-25}$$

其中，$N(t_k, C_i)$是训练集中含有特征 t_k 且属于 C_i 类的样本数，$N(C_i)$ 为训练集中 C_i 类样本的数目，M 表示类别的数量。

2. 多项式模型（Multinomial Model）

若目标 T 采用 TF 向量表示法[129]，即模式向量 $\boldsymbol{V}$ 的分量为相应特征项在该目标中出现的频度。则目标 T 属于 C_i 类的概率为

$$P(C_i|T)=\frac{P(C_i)\prod_{t_k\in V}P(t_k|C_i)^{TF(t_k,T)}}{\sum_i\left[P(C_i)\prod_{t_k\in V}P(t_k|C_i)^{TF(t_k,T)}\right]} \tag{3-26}$$

其中，$TF(t_k, T)$是目标 T 中特征 t_k出现的频度，$P(t_k|C_i)$ 是对在 C_i 类目标中

特征 t_k 出现的条件概率的拉普拉斯估计：

$$P(t_k | C_i) = \frac{1 + TF(t_k, C_i)}{|V| + \sum_{t_k \in V} TF(t_k, C_i)} \tag{3-27}$$

这里，$TF(t_k, C_i)$是 C_i 类目标中特征 t_k 出现的频度，$|\boldsymbol{V}|$为特征分量的总数，即目标表示中所包含的不同视觉单词的总数目。

朴素贝叶斯模型所需估计的参数很少，对缺失数据不太敏感，算法也比较简单。它可以在线性时间内学习完所有的训练集，并渐近地更新其参数，数据到达的顺序和分类错误均不影响分类器的学习过程。理论上，朴素贝叶斯分类器与其他分类方法相比具有最小的误差率。但是该模型在分类识别中假设特征项之间相互独立，而这个假设在实际应用中往往是不成立的，这给朴素贝叶斯分类器的正确分类带来了一定影响。因此，近年来大量的研究工作致力于改进朴素贝叶斯分类器，主要集中在选择特征子集和放松独立性假设在两个方面。

3.5.3　基于 BP 神经网络的分类器

人工神经网络是在对人脑神经网络的基本认识的基础上，用数理方法从信息处理的角度对人脑神经网络进行抽象，建立的某种简化模型[113]。其中，反向传播网络（Error Back Propagation Neural Network）是迄今为止应用最广泛的一种神经网络，它是使用 BP 算法进行学习的多级非循环网络。BP 算法在于利用输出层的误差来估计输出层的直接前导层的误差，再用这个误差估计更前一层的误差，这样就形成了将输出端表现出的误差沿着与输入信号相反的方向逐级向网络的输入端传递的过程。BP 算法结束了多层网络没有训练算法的历史，并被认为是多级网络系统的训练方法，它有很强的数学基础，故其连接权的修改是令人信服的。

1. 三层 BP 网络设计

BP 网络的结构设计主要是解决设几个隐含层和每层设几个节点的问题。对于这类问题，不存在通用性的理论指导，但神经网络的设计者们通过大量的实践已经积累了不少经验。因为已有结果表明一层隐含层已经足够近似任何连续函数，故图像目标识别系统常常采用三层 BP 神经网络。第一层输入层 PE（处理单元）的数量通常由应用来决定，它可以等于特征向量的维数；第二层隐含层的 PE 数量则是设计时需要选择的，由于不知道确定神经网络内部层次中间节点数目的规则，因此这个数目一般基于以前的经验或任意指定并通过检验来

完善。

如图 3-3 所示，特征向量的维数 N 即为 BP 网络的输入层节点数；中间隐含层的神经元数目确定为$(N+N_o)/2$（输入和输出层神经元的平均数）。通常为了减少过度训练的危险，需要将这个数量尽量减少，但是太少又会使网络无法收敛到一个对复杂特征空间恰当的划分，所以网络收敛后，一般可以减少 PE 的数量再进行训练会得到更好的效果。输出层的节点数与模式类的数目一致，从上到下的 N_o个节点代表各个类别 ω_j（$j=1$，2，3，4）。在设定网络结构后，我们对整个网络使用同样形式的"S"激活函数，权值被初始化为带有零均值的小随机数，然后使用模型投影图的相应模式向量对网络进行训练。输出节点在训练期间是受到监控的。对类 ω_i的所有训练模式，与所求类一致的输出节点必须为高（≥0.95），而同时，所有其他节点必须为低（≤0.05）。

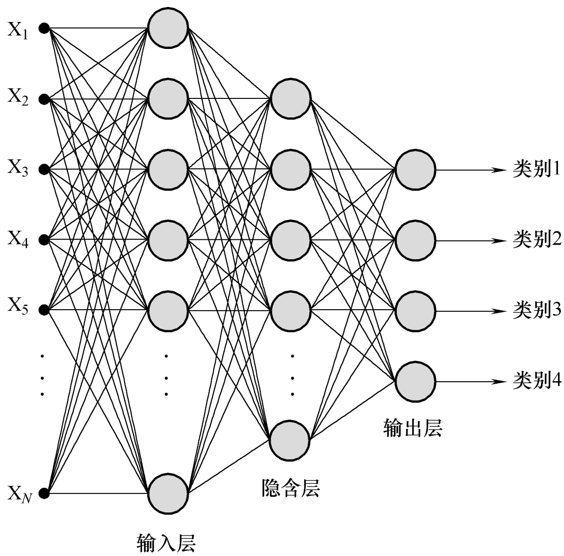


图 3-3　用于目标分类的三层 BP 网络

2. 训练方法的改进

BP 网络接受样本的顺序会对训练的结果有较大的影响。比较而言，它更"偏爱"较后出现的样本：如果每次都按照$(\boldsymbol{x}_1, \boldsymbol{y}_1)$，$(\boldsymbol{x}_2, \boldsymbol{y}_2)$，…，$(\boldsymbol{x}_s, \boldsymbol{y}_s)$所给定的顺序进行训练（$s$ 为样本数目，$\boldsymbol{x}_i$ 为输入向量，$\boldsymbol{y}_i$ 为输出向量，$i=1, 2, \cdots, s$），在网络学习完成投入运行后，对于与该样本序列较后的样本较接近的输入，网络所给出的输出的精度将明显高于与样本序列较前的样本较接近的输入对应的输出的精度。

实际上，按照这种方法进行训练，有时甚至会引起训练过程的严重抖动，更严重的，它可能使网络难以达到用户要求的训练精度。这是因为排在较前的样本对网络的部分影响被排在较后的样本的影响掩盖掉了，从而使排在较后的样本对最终结果的影响就要比排在较前的样本的影响大。这表明，虽然知识的分布表示原理告诉我们，信息的局部破坏不会对原信息产生致命的影响，但是这个被允许的破坏是非常有限的。此外，算法在根据后来的样本修改网络的连接矩阵时，进行的是全面修改，这使得“信息的破坏”也变得不再是局部的。这正是 BP 网络在遇到新内容时，必须重新对整个样本集进行学习的主要原因。

因此，在训练网络的时候，本书采用随机抽取的方法选取样本。在一轮训练过程中，每次都从 s 个样本中随机选取一个样本进行训练，直到所有 s 个样本全部都被选取过。系统进行训练之后，使用在训练阶段中设定的参量对模式进行分类。在标准操作中，所有反馈路径是不连通的。任何输入模式允许通过不同层进行传播，并且模式被划归为高的节点输出所属的类。此时，其他所有节点输出为低。如果被标记为高的节点不止一个，或没有节点输出被标记为高，则可选的做法是，声明进行了错误的分类或简单地将模式划归输出节点的类并赋予最大值。

3.5.4　基于支持向量机的分类器

支持向量机是 Vapnik 及其合作者[130]根据结构风险最小化原则提出的一种在高维特征空间使用线性函数假设空间的学习系统。支持向量机是机器学习领域若干标准技术的集大成者。它集成了最大间隔超平面、Mercer 核、凸二次规划、稀疏解和松弛变量等多项技术。在若干挑战性的应用中，获得了目前为止最好的性能。

1. 线性分类

两类模式（正类和负类）的识别通常用一个实数函数 f：$X\subseteq R^n\rightarrow R$（$n$ 为输入维数，R 为实数）。通过执行如下操作：当 $f(x)\geqslant 0$ 时，将输入 $x=(x_1, x_2, \cdots, x_n)'$ 标记为正类，否则，将其标记为负类。当 $f(x)(x\in X)$ 是线性函数时，$f(x)$ 可以写成如下形式：

$$f(x)=\langle w\cdot x\rangle+b=\sum_{i=1}^{n} w_i x_i+b \tag{3-28}$$

式中，$(w, b)\in R^n\times R$，是控制函数的参数，决策规则由函数 sgn（$f(x)$）给出，

通常 sgn(0) =1，学习意味着要从数据中获得这些参数；“·”是向量点积。

该类方法的几何解释是，方程式$\langle w \cdot x \rangle + b = 0$定义的超平面将输入空间$X$分成两个部分。如图 3-4 所示，黑斜线表示超平面，w是超平面的法线方向。当b值变化时，超平面平行于自身移动。因此，如果表达R^n中所有可能的超平面，一般要包括$n+1$个可调参数的表达式。

如果训练数据可以无误差地被划分，那么，以最大间隔分开数据的超平面称为最优超平面，如图 3-5 所示。

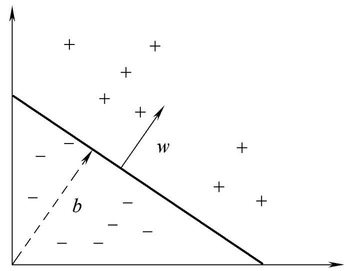


图 3-4　二维训练集的分开超平面(w, b)

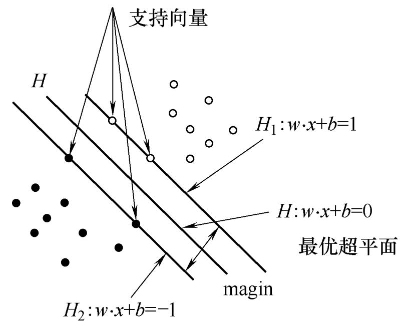


图 3-5　最优超平面

对于多个模式类的分类问题，输出域是$Y = \{1, 2, \cdots, m\}$。线性学习器推广到$m(m \in N, m \geqslant 2)$类是很直接的：给每个类关联一个超平面，然后，将待分类的数据点赋予超平面离其最远的那一个类。输入空间分为m个简单相连的凸区域。

2. 线性不可分

对于非线性问题，可以把样本x映射到某个高维特征空间，在高维特征空间中使用线性学习器。因此，考虑的假设集是这种类型的函数：

$$f(x) = \sum_{i=1}^{N} w_i \phi_i(x) + b \tag{3-29}$$

式中，ϕ：$X \to F$是从输入空间到某个特征空间的映射，如图 3-6 所示。也就是说，建立非线性分类器需要分两步：首先使用一个非线性映射函数将数据变换到一个特征空间F，然后在这个特征空间上使用线性分类器。

线性分类器的一个重要性质是可以表示成对偶形式，这意味着假设可以表达为训练点的线性组合，因此，决策规则（分类函数）可以用测试点和训练点的内积表示：

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i \langle \phi(x_i) \cdot \phi(x) \rangle + b \tag{3-30}$$

式中，l 是样本数目；α_i 是个正值导数，可通过学习获得；y_i 为类别标记。如果有一种方法可以在特征空间中直接计算内积 $\langle\phi(x_i)\cdot\phi(x)\rangle$，就像在原始输入点的函数中一样，那么，就有可能将两个步骤融合到一起建立一个非线性分类器。这样，在高维空间内实际上只需要进行内积运算，而这种内积运算是可以利用原空间中的函数实现的，我们甚至没有必要知道变换的形式。这种直接计算的方法称为核（Kernel）函数方法。

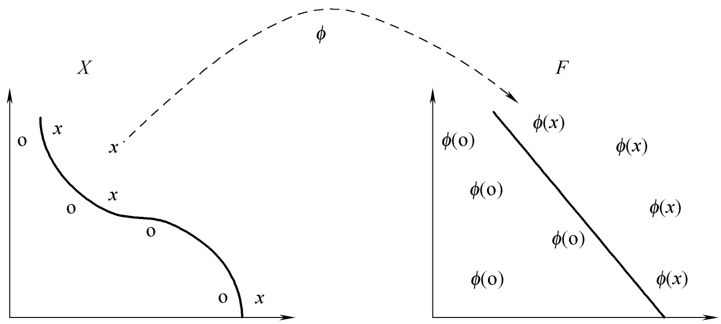


图 3-6　简化分类任务的特征映射

3. 构造核函数

定义 3-1　核是一个函数 K，对于所有 x，$z\in X$，满足

$$K(x,\ z)=\langle\phi(x)\cdot\phi(z)\rangle \tag{3-31}$$

这里 ϕ 是从 X 到特征（内积）空间 F 的映射。

一旦有了核函数，决策规则就可以通过对核函数的 l 次计算得到

$$f(x)=\sum_{i=1}^{l}\alpha_i y_i K(x_i,x)+b \tag{3-32}$$

那么，这种方法的关键就是如何找到一个可以高效计算的核函数。

核函数要适合某个特征空间必须是对称的，即

$$K(x,\ z)=\langle\phi(x)\cdot\phi(z)\rangle=\langle\phi(z)\cdot\phi(x)\rangle=K(z,x) \tag{3-33}$$

并且，满足下面的不等式：

$$\begin{aligned}K(x,\ z)^2&=\langle\phi(x)\cdot\phi(z)\rangle^2\leqslant\|\phi(x)\|^2\|\phi(z)\|^2\\&=\langle\phi(x)\cdot\phi(x)\rangle\langle\phi(z)\cdot\phi(z)\rangle=K(x,\ x)K(z,\ z)\end{aligned} \tag{3-34}$$

其中，$\|\cdot\|$ 是欧氏模函数。但是，这些条件对于保证特征空间的存在时不充分的，还必须满足 Mercer 定理的条件，对 X 的任意有限子集，相应的矩阵式半正定的。也就是说，令 X 是有限输入空间，$K(x,\ z)$ 是 X 上的对称函数。那么，$K(x,\ z)$ 是核函数的充分必要条件是矩阵

$$K = (K(x_i, x_j))_{i,j=1}^{n} \tag{3-35}$$

是半正定的（即特征值非负）。

根据泛函的有关理论，只要一种核函数满足 Mercer 条件，它就对应某一空间中的内积。目前 SVM 常用的核函数有

线性核： $$K(x, z) = \langle x \cdot z \rangle \tag{3-36}$$

多项式核：$K(x, z) = (\langle x \cdot z \rangle + c)^d$，其中 $c \geq 0$，$d \in N$；当 $c = 0$ 时，称为齐次多项式核，当 $c > 0$ 时，称为非齐次多项式核。 (3-37)

高斯（径向基）核： $$K(x, z) = \exp(-\|x - z\|^2 / 2\sigma^2), \sigma > 0 \tag{3-38}$$

Sigmoid 核： $$K(x, z) = \tanh(v\langle x \cdot z \rangle + c) \tag{3-39}$$

3.6 本章小结

概括来说，识别过程就是通过找出描述并区分数据类或概念的模型（或函数），以便能够使用模型预测那些未知标记的对象类。正如很多科研人员的共识，图像目标识别作为机器智能的重要方面，仍处于实践发展的初级阶段，面对计算机科学中图像工程和机器视觉里的许多关键问题，需要借鉴认知科学领域中人工智能和模式识别的许多经典方法。

本章首先从决策理论和结构判别两个方面简述了模式识别的主要技术和对应的模式形式；然后探讨了目标匹配的基本方法和模式匹配的相似度度量问题，即模式向量的距离问题；接着对目标分类器的研究现状进行了综述，主要围绕着分类器的种类及其性能评估方法；最后，详细介绍和比较了几种典型的图像目标分类器的原理与特点。正如第 2 章所述，目标识别和目标分割是紧密相关的，实际应用中也常常需要数据驱动和理论驱动相结合，在整个认知过程中分割和识别交替进行，这也是图像工程和机器视觉领域的一个发展趋势。

第 4 章　图像目标的局部特征提取

科学中像制造业一样，更换工具是一种浪费，只有在不得已时才会这么做。危机的意义就在于，它指出更换工具的时代已经到来了。

——托马斯 · 库恩（1922—1996）

4.1　引言

传统的特征提取方法大都将目标作为一个整体，从大量包含目标的图片集中学习并提取整体特征，如面积、周长、不变矩和傅里叶描绘子等，并采用统计分类技术进行目标分类。这种识别方法有以下几个缺点：对于结构复杂的图像，识别效果受到图像分割精度的制约；需要学习大量的数据以及较长的训练时间；由于没有捕捉到图像中物体的局部信息，当目标的形状发生较大变化时，比如目标被局部遮挡，就会导致整体特征的突然变化，对于目标识别是非常不利的[131]。

大量研究表明，人类视觉系统可以将物体分解为许多有意义的小块，并通过这些局部的信息进行目标的辨识[132]。这使得采用局部特征技术在复杂背景下的目标识别上有着越来越广泛的应用。局部特征目前还没有一个统一的定义，

它的提出主要是相对整体特征而言，用局部特征对图像进行描述时可以得到图像中物体的局部信息。在图像内容复杂、噪声干扰较大、存在局部遮挡、目标姿态发生较大变化等情况下，利用局部信息进行目标识别是非常有效的。

局部特征提取一般包括特征区域检测和特征区域描述两部分内容，从广义上讲，还包含对特征空间的进一步优化。与分类器设计相比，局部特征提取更加依赖于具体问题和相应领域的知识。而且从实用的角度来说，大多数局部特征都要求对亮度、尺度、平移和旋转具有一定的不变性。

近几年来，对局部特征的研究非常活跃，新的方法不断涌现。本章在对国内外众多研究成果深入探讨之后，根据后续实验的需要，选用并改进了一些特征区域检测算法和特征区域描述算子，为不同情况下的目标识别提供了合适的局部特征。

4.2 特征区域的稀疏选取算法

4.2.1 特征区域检测的研究现状

目前，常用的特征区域检测方法可以分为三类，分别是密集选取、稀疏选取和其他选取方法。从本质上看，所有的这些方法都是建立在对图像像素遍历的基础之上的。

1. 密集选取方法

这种方法的研究者普遍持有这样一种观点：在模式识别的低层处理中，所有图像区域都有一定的作用，丢失任何细节都可能对最终效果产生很大的影响。Ohba 和 Ikeuchi[133,134]提出将图像密集地分为互不重叠的特征窗（Eigen Windows），每个特征窗都当作一个局部特征区域；Jurie[78]以整幅图像的每一个像素点为中心，选取周围的区域作为局部特征区域；Dalal[135]和 Zhu[136]采用在检测窗口的每个像素位置、不同尺度下提取大量的特征区域，以供进一步应用。

密集选取方法在滑动窗口模型中应用较多，其优点就是基本没有丢失图像的细节，可以得到非常丰富的局部特征。但是其中很大一部分特征区域信息量过小，对后期的识别没有作用甚至起到干扰作用，加重了下一步特征优化工作的负担。

2. 稀疏选取方法

这种方法都是通过特征检测，选取具有显著特点的图像区域作为局部特征。

检测算子一般可以分为基于形状（Shape-based）的检测算子和基于外观（Appearance-based）的检测算子两类。

基于形状的检测算子是根据图像的形状特征（如边界、直线、弧线等）来确定特征区域的位置。主要应用于外形区分度明显的目标识别，如各种刚性的、无关节的物体。Gool[137]利用图像的边缘信息对图像进行分析和理解，构造了线矩特征，作为一种局部信息量，它受到平移、旋转和尺度变化的影响较小；Belongie[138]围绕着梯度算子检测出的边缘点，提出了 SC（Shape Context）特征，描述子的维度为 36；Berg[139]结合边缘方向能量与高斯核函数，得到了一种 204 维的局部特征，命名为 GB（Geometic Blur）；Fergus[140]用 Canny 算法检测图像的边缘，选择边缘点周围的区域作为特征区域。

基于外观的检测算子是在图像的灰度模式下，搜寻具有某种稳定性和不变性的特征点或关键区域。Beaudet[141]通过对图像函数二阶导数的泰勒展开，得到了具有旋转不变性的 Hessian 矩阵，可以直接对灰度图像进行操作提取特征点；Harris 等人[142]受到了信号处理中自相关函数的启发，提出了 Harris 算法，也称为 Plessey 算法，这种算法是通过自相关矩阵来检测特征点的；随后，Mikolajczyk 和 Schmid 等人结合拉普拉斯和高斯变换对 Hessian 和 Harris 算法进行了改进，提出了 Harris-Laplace[143]，Hessian-Laplace[144]，Harris affine [144]，Hessian affine[145]四种检测算子；Lowe[146]提出的高斯差分（Difference of Gaussian，DoG）算子是在尺度空间寻找极值点，结果比较稳定，抗噪能力较强；Kadir 等人[147]提出的 SalReg（Salient Regions）算子，利用亮度直方图在尺度空间计算局部最大熵，将其所对应的圆形区域定义为特征区域；Matas 等人[148]结合分水岭算法和阈值思想提出了 MSER（Maximally Stable Extremal Regions）算法，检测出的灰度值居中的稳定区域。

稀疏选取法检测出的特征区域数量一般在 200～3000，其主要优点是简洁、紧致，图像的关键点远少于图像的像素，使得后面的识别过程能大大加速。但很多特征区域检测算法往往和图像的特性相关，应用到通用目标识别时，可能会有一定的局限。

3. 其他选取方法

Nowak[149]在研究向量空间模型的取样策略时发现，当训练集的样本足够多时，随机取样法能达到和某些稀疏取样相近甚至更好的结果。Moosmann 等研究者[32]提出了使用显著性映射在分类过程中动态选取图像块的方法。

三类特征区域检测方法都是建立在扫描、分析整幅输入图像的基础之上的，

不同的是：密集选取方法在滑动窗口模型中应用较多，其优点就是基本没有丢失图像的细节，可以得到非常丰富的局部特征，但是其中很大一部分特征区域信息量过小，对后期的识别没有作用甚至起到干扰作用；随机选取等方法需要的训练集样本数量较大，这本身就加重了后面分类识别的负担；稀疏选取目前被广泛应用于各种目标识别系统，而且可供选用的算子不断涌现，但每个算子的效果往往和目标以及背景的特性有很大的关联，所以如何选择合适的检测算子是进行目标识别的关键。

4.2.2　高斯差分检测算子

近几年，高斯差分（Difference of Gaussian，DoG）、SalReg（Salient Regions）、MSER（Maximally Stable Extremal Regions）算法的相继出现，代表着基于外观的检测算子开始广泛应用于机器视觉领域。高斯差分算子是在多尺度空间中寻找稳定有效的特征区域。Koendetink 和 Lindeberg 等人[150-152]证明了高斯卷积核是实现尺度变换的唯一线性核，所以，一幅二维图像 $I(x, y)$的尺度空间定义为

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{4-1}$$

式中，符号 $*$ 表示卷积，(x, y)代表图像中像素的位置，而尺度可变高斯函数为

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2} \tag{4-2}$$

利用不同尺度的高斯差分算子与图像进行卷积运算，可以求取尺度空间极值，计算公式如下：

$$\begin{aligned}D(x, y, \sigma)&=[G(x, y, k\sigma)-G(x, y, \sigma)]*I(x, y)\\&=L(x, y, k\sigma)-L(x, y, \sigma)\end{aligned} \tag{4-3}$$

其中，L 代表了图像的尺度空间，k 是常数，一般取 $k=\sqrt{2}$。这里选择高斯差分函数的原因主要有两个：一是其计算效率较高；二是它可以作为尺度归一化的高斯拉普拉斯函数（Laplacian of Gaussian，LoG）——$\sigma^2\nabla^2G$ 的一种近似[152]，如图 4-1 所示。

通过与其他特征提取算子（如 Harris、Hessian 算子）的实验比较，Mikolajczyk 等人[154]发现基于 $\sigma^2\nabla^2G$ 的极大值和极小值能够产生更为稳定的局部特征。$D(x, y, \sigma)$与 $\sigma^2\nabla^2G$ 的关系可以从如下公式推导得到：

$$\frac{\partial G}{\partial\sigma}=\sigma\nabla^2G \tag{4-4}$$

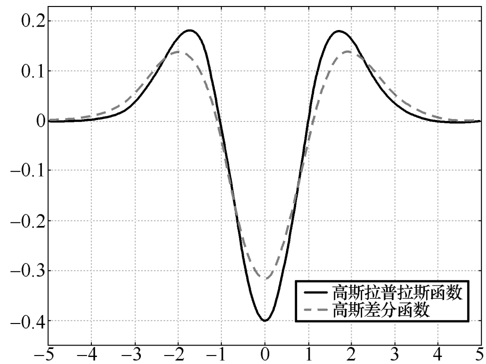


图 4-1　高斯拉普拉斯函数与高斯差分函数

利用差分近似替代微分，则有

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma} \tag{4-5}$$

因此，有

$$G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2 \nabla^2 G \tag{4-6}$$

其中 $k-1$ 是个常数，并不影响极值点位置的求取。

如图 4-2 所示，Lowe 等人[153]提出了一种构造 $D(x, y, \sigma)$ 的有效方法。左侧是不同尺度空间中的图像金字塔，右侧显示了将每层金字塔中相邻图像相减所生成的高斯差分图像的结果。

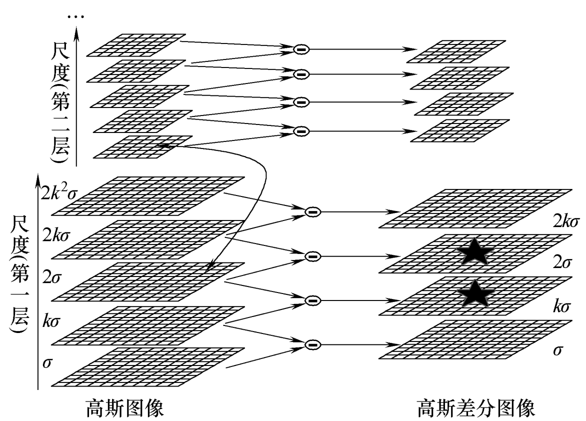


图 4-2　高斯金字塔和高斯差分金字塔的构造示意图

限于篇幅，图 4-2 只给出了第一层和第二层高斯差分图像的计算。在实际应用中，高斯金字塔一般选择 4 层，每层有 5 幅一组的尺度图像。在目前常用的

设计方案中，第一层的第一幅图像是放大 2 倍的原始图像，其目的是为了得到更多的特征点。

图 4-3 是利用一幅关于爱因斯坦的图像构造高斯金字塔和高斯差分金字塔的示例。图 4-3b 所示的高斯差分图像是通过图 4-3a 金字塔中对应层上的相邻图像相减而得到的。



a)



b)

图 4-3　两种图像金字塔的示例

a）高斯金字塔　b）高斯差分金字塔

图 4-4 所示为如何从高斯差分金字塔的分层结构中提取出图像的极值点作为候选的特征点，就是将每个检测点与其相邻点（图像域和尺度域）进行逐个比较，得到的局部极值位置即为该特征点所处的位置和对应的尺度。如图 4-2 中右图的五角星符号所标识，由于需要与相邻尺度的点进行比较，所以在每层高斯差分金字塔的一组图像中只能检测到两个尺度的极值点。

由于 DoG 算子对噪声和边缘较为敏感，因此，在上面 DoG 尺度空间中检测到的局部极值点还需要经过进一步的检验才能精确定位为特征点。通过拟合三

维二次函数可以较精确地计算特征点的位置和尺度，同时还可以去除对比度较低的特征点以及稳定性较差的边缘响应点[146]。

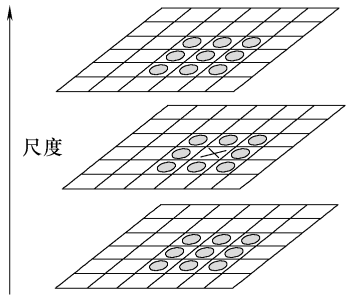


图 4-4 尺度空间的极值检测

（来源：Lowe，2004）

获取特征点处的拟合函数为

$$D(X) = D + \frac{\partial D^T}{\partial X}X + \frac{1}{2}X^T\frac{\partial^2 D}{\partial X^2}X \tag{4-7}$$

求导并让方程等于零，可以得到极值点：

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2}\frac{\partial D}{\partial X} \tag{4-8}$$

对应极值点，方程的值为

$$D(\hat{X}) = D + \frac{1}{2}\frac{\partial D^T}{\partial X}\hat{X} \tag{4-9}$$

如果 $|D(\hat{X})| \leqslant 0.03$，则视为对比度较低的候选特征点，并予以剔除。

因为 DoG 算子会产生较强的边缘响应，所以需要对这些不稳定的点进行检测。首先获取该点处的 Hessian 矩阵：

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{4-10}$$

H 的特征值 α 和 β 代表 x 和 y 方向的梯度：

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta$$
$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \tag{4-11}$$

$Tr(H)$和 $Det(H)$分别表示矩阵 H 的迹与行列式。假设 α 是最大的特征值，β 是较小的特征值，令 $\alpha = r\beta$，则

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \tag{4-12}$$

式（4-12）的值越大也就是两个特征值之比越大，这就说明在某一个方向上的梯度值越大，同时另一个方向上的梯度值越小，这种情况恰恰符合边缘响应的条件。一般取 $r = 10$，并检测是否符合以下条件，就可以剔除边缘响应点：

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \tag{4-13}$$

4.2.3 边缘点检测算子

基于形状的检测算子一般都是将边缘点作为特征点，从而进行特征描述的。Canny[155] 提出了评价边缘检测算法性能优良的三个指标：

1）高的信噪比；

2）精确的定位性能；

3）对单一边缘响应是唯一的。

Canny 算子首次将上述判据用数学的形式表达出来，然后采用最优化数值方法，得到对应给定边缘模型的最佳边缘检测模板。对于二维图像，需要使用若干方向的模板分别对图像进行卷积处理，再取最可能的边缘方向。Canny 的分析是针对一维边缘中的阶跃型边缘，Canny 推导出的最优边缘检测器的形状与高斯函数的一阶导数类似，利用二维高斯函数的对称性和可分解性，可以很容易计算高斯函数在任意方向上的方向导数与图像的卷积。因此，在实际运用中可以选取高斯函数的一阶导数作为阶跃边缘的次最优检验算子。

设图像 $f(x, y)$，二维高斯函数如式（2-2）所示，由卷积求导性质可知，$G*f(x, y)$的梯度为

$$\nabla[G*f(x,y)] = \nabla G*f(x,y) \tag{4-14}$$

梯度的模值，即 $G*f(x, y)$的最大方向导数为

$$|\nabla[G*f(x,y)]| = \left[\left(\frac{\partial G}{\partial x}*f\right)^2 + \left(\frac{\partial G}{\partial y}*f\right)^2\right]^{1/2} \tag{4-15}$$

以其作为边界强度。

梯度的单位方向矢量

$$\vec{\boldsymbol{n}} = \frac{\nabla G*f(x,y)}{|\nabla G*f(x,y)|} = (\cos\alpha, \sin\alpha)' \tag{4-16}$$

式中

$$\sin\alpha = \frac{\frac{\partial G}{\partial y}*f}{|\nabla G*f|}, \cos\alpha = \frac{\frac{\partial G}{\partial x}*f}{|\nabla G*f|} \tag{4-17}$$

$\vec{\boldsymbol{n}}$ 或 α 给出了边界的法线方向。

以上为 Canny 二维最优阶跃边缘检测算子的数学推导。在实际应用中，可以将原始模板截断到有限尺寸 N，为了提高运算速度，可以将 ∇G 的二维卷积模板分解为两个一维卷积模板。

根据 Canny 边缘的提取原则。当一个像素满足以下三个条件时，则被认为是图像的边缘点：

1）该点的边缘强度大于沿该点梯度方向（这里指正反向）上的两个相邻像素点的边缘强度——主要作用是准确定位并控制边缘宽度为一个像素点。

2）与该点梯度方向上相邻两点的梯度方向之差小于 45°——给出光滑性约

束，克服随机因素的影响。

3）以该点为中心的 3 × 3 邻域中的边缘强度极大值小于某个阈值——保持边缘强度相对一致，去除噪声产生的伪边缘。

4.3　局部特征的定量描述

4.3.1　特征区域描述的研究现状

在图像中检测出不同的特征区域之后，需要使用一种更适合于计算机进一步处理的形式，对得到的区域像素集进行表示和描述。基本上，表示一个区域包括两种选择：用其外部特性来表示区域（如区域的边界）；用其内部特性来表示区域（如组成区域的像素）[8]。显然，一般局部特征区域的外部特性不具有区分性，只能通过其内部特性来表示。

常用的局部特征描述子都是基于选定的表示方式，将特征区域描述为向量的形式，又称特征向量。这些特征描述子一方面要充分体现出不同目标的差异，又要易于计算局部特征之间的相似度，还要对背景噪声和目标姿态的变化具有鲁棒性。Mikolajczyk[156]将局部特征描述子从技术应用角度分为四大类：基于分布的描述子、基于空间频率技术的描述子、差分描述子和其他描述子。

1. 基于分布的描述子

这类描述子主要利用直方图来描述不同的外观或形状特征。一种最简单的描述子就是用灰度直方图来描述区域中像素点的强度分布；在亮度变化的情况下，使用区域灰度级直方图的统计矩[8]效果更好，但它的应用局限于对纹理图像的描述；SI（Spin Image）[157]通过对围绕着区域中心点的 5 个环分别统计灰度值，使得描述子对亮度变化、旋转变化不敏感；Lowe[146]提出的 SIFT（Scale Invariant Feature Transform）描述子是通过 DoG 检测子和梯度方向直方图获得每个关键点的位置、尺度和方向信息，并利用坐标轴旋转、多种子点联合描述、向量长度归一化等技术消除了旋转、光照和尺度变化等因素的影响，该描述子适用范围广、运算速度快、鲁棒性强；GH（Geometric Histogram）[158]和 SC[138]描述子的主要思想与 SIFT 描述子类似，只是它们描述的是区域内边缘的分布，主要应用于边缘特征比较明显、稳定的图像；PCA-SIFT[159]和 GLOH（Gradient Location Orientation Histograms）[156]描述子都是对 SIFT 描述子的扩展，它们在区域和梯度方向上采用了不同的描述精度，并用主分量分析对特征向量进行降维处理，

进一步增强了描述子的鲁棒性和区分度。

2. 基于空间频率技术的描述子

这类方法的优势在于，通过用频域技术对图像进行描述和处理，可以充分利用频率成分和图像外观之间的对应关系。但最初的傅里叶变换是将图像信号转化为无限域的基函数，而且像素点之间的空间关系是不明确的，这极不适用于局部特征。Gabor 滤波器和小波变换则克服了上述缺陷，被广泛应用于纹理图像的分类和识别中。Papageorgiou[160]、Mohan[161]和 Viola[3]等人将图像由空间域映射到频域，采用类似于 Haar 小波的频谱方法表示图像区域，结合支持向量机和核方法，实现了行人、人脸和汽车等目标的检测与识别。

3. 差分描述子

一系列的图像导数也可以用来描述一个点附近的区域特征。Koenderink 和 Doorn[162]就提出了用差分计算来获取导数的近似，并得到了 local jet 描述子；此后，Florack 等人[163]又改进了该描述子，使其具有旋转不变性；Freeman 和 Adelson[164]提出的导向滤波器（Steerable Filters）是对 local jet 的进一步完善，它通过与高斯导数卷积并调整导数沿着梯度方向，使得该描述子适用于旋转和光照变化的图像；复数滤波器（Complex Filters）是利用方程 $K(x, y, \theta)=f(x, y)\exp(i\theta)$的求导结果对区域进行描述的，其中 θ 是方向，而 $f(x, y)$ 的形式要根据具体情况而定，Baumberg[165]用的是高斯导数，Schaffalitzky 和 Zisserman[166]则用多项式。

4. 其他描述子

Gool[137]提出的广义不变矩是指物体图像经过平移、旋转以及比例变换仍保持不变的矩特征量。不变矩描述了一个区域内的形状和亮度分布，它的特征维数较少，对彩色图像的每个颜色通道的计算结果都很稳定，但高阶矩对几何失真和光亮度失真比较敏感。基于人类对纹理的视觉感知的心理学的研究，Tamura 等人[167]提出了纹理特征的表达，它用以描述特征区域的六个分量分别是对比度、方向度、粗糙度、线像度、规整度和粗略度，这种局部特征常用于图像检索领域，对纹理图像的识别效果比较好。

随着技术的进步，不断有新的描述子出现，但每种描述子都有一定的适用范围，而且其性能与特征区域检测方法没有必然的联系。总体看来，GLOH 和 SIFT 描述子应用比较广泛，性能比较稳定；SC 描述子在形状特征明显的目标识别中效果很好，但在纹理图像和非刚性目标的识别中效果不佳；在低维描述子中，不变矩和导向滤波器的性能要略胜一筹。

4.3.2　基于梯度分布的描述子

Lowe 提出的 SIFT 描述子对后来的许多基于梯度分布的特征描述子都产生了深远的影响。例如，GH 和 SC 描述子的主要思想就和 SIFT 描述子类似，只是它们描述的是区域内边缘的分布，主要应用于边缘特征比较明显、稳定的图像；PCA-SIFT 和 GLOH 描述子都是对 SIFT 描述子的扩展，它们在区域和梯度方向上采用了不同的描述精度，并用主分量分析对特征向量进行降维处理，进一步增强了描述子的鲁棒性和区分度。

为了使描述子具有旋转不变性，需要为每一个特征点指定一个方向，从而让局部特征描述子与这个方向因子相关。计算特征点邻域的梯度模值以及梯度方向的公式如下：

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$
$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \tag{4-18}$$

式中，L 为每个特征点各自所在的尺度。在以特征点为中心的邻域窗口内计算像素的梯度方向直方图，直方图的范围是0°～360°，以10°为一个步长，共分为36个方向。如图 4-5 所示，在计算过程中需要一个的高斯权重窗（左图中的圆形），中心处的权值最大，边缘处的权值最小，右图给出了 8 个方向的直方图示例（实际应用中常采用36 个方向）。

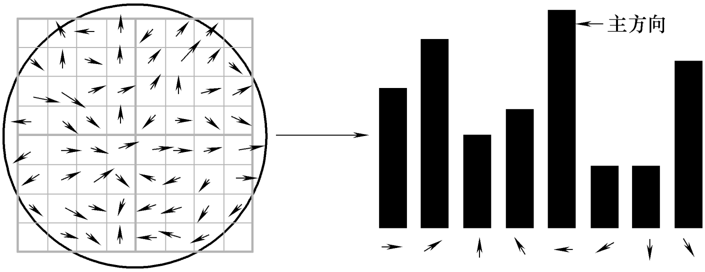


图 4-5　由梯度方向直方图确定主梯度方向

为了增强匹配的稳定性，以梯度方向直方图的最大值作为该特征点的主方向，并选择大于主方向峰值 80% 的方向作为辅方向。虽然在相同位置和尺度可能创建多个特征点但方向不同，且 15% 的特征点被赋予多个方向，这明显提高了特征点的区分性。

如图 4-6 所示，每一个特征点都携带了三种信息——位置、尺度和方向，

由此可以确定一个 SIFT 特征区域，可以将坐标轴旋转为特征点的方向，进而构造出独特性较高的特征描述子，且具有不受尺度、光照、视角变化影响的性质。左图中矩形的中心点表示当前特征点的位置，小箭头的长度代表梯度的幅值，箭头的方向表示梯度的方向。圆形的高斯窗（越靠近中心点，贡献越大）尽量减小那些远离特征区域中心的梯度值影响，这样就避免了微小变化引起的描述子突变。

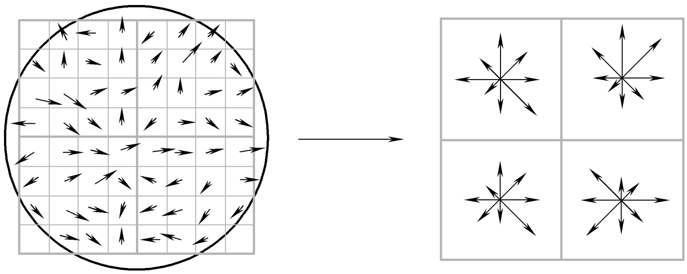


图 4-6　由邻域梯度信息生成特征向量

图 4-6 的右图所示的描述符是基于一个 2 ×2 个梯度方向直方图，Lowe 建议在实际应用中采用 16 个直方图进行描述效果最好。即以特征点为中心取 16 ×16 像素大小的邻域，将此邻域分为 16 个大小为 4 ×4 个像素的子区域，对每个子区域计算 8 方向的梯度方向直方图。根据子区域位置对相应的梯度方向直方图排序，就构成了一个 4 ×4 ×8 =128 维的 SIFT 特征向量。如此一来，该特征描述子就消除了尺度变化、旋转变化等因素的影响，通过向量的长度归一化可以进一步消除光照变化的影响。

GLOH 描述子是对 SIFT 描述思想的改进和发展，首先利用邻域像素的梯度方向分布为每个特征点指定方向参数，并将坐标轴旋转为该方向，以确保旋转不变性。然后在特征点所处的尺度空间（即高斯金字塔的某一层），取其周围的 16 像素 16 像素大小的邻域，用 17 层放射状同心圆来表示，并对每个子区域计算梯度方向直方图（梯度方向分为 16 种）。对 17 个子区域的 16 方向梯度直方图根据位置依次排序，这样就得到一个 17 ×16 =272 维的向量。通过主分量分析（Principal Component Analysis，PCA）进行降维，最终得到一个 128 维的向量，在最大程度保留原始数据的同时大大减少了后续应用的计算时间。

基于梯度分布的特征描述子都可以较稳健的对发生几何形变、退化、受噪声干扰的图像局部特征进行准确的匹配。而且由于这些特征描述子在计算关键

点方向时充分利用了邻域信息，这样在一定程度上可以避免在小运动物体上匹配特征点，因为小运动物体的邻域信息即使去除了尺度和旋转的因素后也仅是具备较少的梯度方向相似性；同时这些特征描述子在计算关键点处的梯度方向时都使用了直方图统计和高斯加权的思想，这就对存在定位偏差的特征点匹配提供了更好的适应性。

4.3.3　线矩特征描述子

如第 2 章 2.3.3 节所述，面矩作为一种全局信息，已经广泛用于完全分割后的目标识别，其具有前面所述的整体特征的优缺点。而针对图像边缘计算的不变矩，我们称之为线矩，作为一种局部特征，它主要利用目标图像的高频信息部分——边缘信息完成对图像的分析与理解[168]。由于通常目标边缘像素点的个数约为目标所有像素点的平方根，所以，用目标边缘像素来表示其形状要比用目标区域内所有的像素点少得多。

设数字图像中的边缘曲线由 N 个离散点组成，即 (x_i, y_i)，$i=1, 2, \cdots, N$，则 $p+q$ 阶线矩的定义为

$$m_{pq} = \sum_{i=1}^{N} x_i^p y_i^q \Delta l_i \tag{4-19}$$

式中，$\Delta l_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$。

相应的 $p+q$ 阶中心矩定义为

$$\mu_{pq} = \sum_{i=1}^{N} (x_i - \bar{x})^p (y_i - \bar{y})^q \Delta l_i \tag{4-20}$$

式中，$\bar{x} = \dfrac{m_{10}}{m_{00}}$，$\bar{y} = \dfrac{m_{01}}{m_{00}}$，点 $(\bar{x}, \bar{y})$ 即为边缘的质心位置。中心矩是与图像的平移无关的。

当对边缘曲线进行尺度变化时，尺度的变化导致曲线长度的变化，相应的变化因子是 k。此时尺度变化后的中心矩成为 $\mu'_{pq} = \mu_{pq} \times k^{p+q+1}$。

用零阶中心矩对其余各阶中心矩进行归一化，可以得到归一化的中心矩为

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{4-21}$$

对曲线来说，要满足尺度不变性，从 $\eta'_{pq} = \eta_{pq}$ 可推出

$$\gamma = p + q + 1 \tag{4-22}$$

为了使矩描述子与平移、大小、旋转等因素无关，利用 2 阶和 3 阶归一化中心矩可以导出下面 7 个矩不变式：

$$
\begin{aligned}
\varphi_1 &= \eta_{20} + \eta_{02} \\
\varphi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\varphi_3 &= (\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2 \\
\varphi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \\
\varphi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] \\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] \\
\varphi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] \\
&\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21}) \\
\varphi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2] \\
&\quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2]
\end{aligned}
\tag{4-23}
$$

由于 $\varphi_5^2 + \varphi_7^2 = \varphi_3\varphi_4^3$，所以上面的 7 个矩不变式只有 6 个是独立的。

在不变矩的实际计算过程中，如医学图像，其不变矩数值分布范围非常大。而在识别过程中，如果不变矩特征值愈小，对识别结果的贡献就愈小；如果不变矩特征值愈大，对识别结果的贡献就愈大。为此，对 7 个不变矩进行如下修正，以调整其取值范围：

$$
t_i = |\lg(|\varphi_i|)| \qquad i = 1, 2, \cdots, 7 \tag{4-24}
$$

式中所进行的修正变换应综合考虑不变矩特征的大小及后续识别的结构特点。

4.4　角点的检测算法

对于特征点目前尚无严格的定义，在一些文献中又被称为兴趣点、显著点、关键点，有时也和角点的概念混用[169]。以点的位置来表示的点特征是一种最简单的图像特征。事实上，特征点既是一个点的位置辨识，同时也说明它的局部邻域具有一定的模式特征。在参考文献［170］中作者将特征点分为两类：广义特征点和狭义特征点。上述特征区域检测中的特征点就是广义特征点，它本身的位置不具备特征意义，只代表满足一定特征条件的特征区域的位置，这种特征可以不是物理意义上的特征，只要满足一定的数学描述就可以。因此，从本质上说，广义特征点可以认为是一个抽象的特征区域，它的属性就是特征区域具备的属性。而狭义特征点的位置本身具有常规的属性意义，比如角点、交叉点等等。

对角点不同的理解产生了关于角点的不同定义，如图像中具有周围灰度变化剧烈特征的点；图像边界上具有曲率足够高的点；图像中具有最大偏转角和

偏差的点；两条边界以一定的角度相交的地方、边界方向发生剧变的地方以及图像灰度梯度方向变化较大的地方等。在上述思想的指导下产生了许多角点检测算法，其中，直线投影法和 SUSAN 检测法是目前最为常用的两种。

4.4.1　直线投影检测算法

直线投影法是一种基于边界的角点检测算法，其核心思想就是把角点定义在目标的轮廓线上，先分割图像，抽取目标边界的 Freeman 链码，将方向改变程度较大的点标记为角点。

设 L 为目标区域边界，其局部连续链码可表示为

$$L_j^s = \{a_{j-s+1}a_{j-s+2}\cdots a_j\} \tag{4-25}$$

式中，s 为链码的环数；j 为链码的终点；a_i 为点 $i-1$ 到 i 的方向码（$i=j-s+1$，$j-s+2$，…，j）。L_j^s 在 x 和 y 方向的投影，即在链码 7 和 1 的方向的投影为

$$x_j^s = \sum_{i=j-s+1}^{j} a_{i7}, \qquad y_j^s = \sum_{i=j-s+1}^{j} a_{i1} \tag{4-26}$$

式中，a_{i7}，a_{i1} 的值由方向码 a_i 的值确定，见表 4-1。

表 4-1　a_i 与 a_{i7}、a_{i1} 的关系

a_i	0	1	2	3	4	5	6	7
a_{i7}	$\sqrt{2}/2$	0	$\sqrt{2}/2$	−1	$\sqrt{2}/2$	0	$\sqrt{2}/2$	1
a_{i1}	$\sqrt{2}/2$	1	$\sqrt{2}/2$	0	$\sqrt{2}/2$	−1	$\sqrt{2}/2$	0

当 s 值较小时，可以将其看成直线，即有链码的向量表示形式为

$$\vec{\boldsymbol{L}}_j^s = x_j^s\vec{\boldsymbol{i}} + y_j^s\vec{\boldsymbol{j}} \tag{4-27}$$

那么，其长度可以表示为

$$|\vec{\boldsymbol{L}}_j^s| = \sqrt{(x_j^s)^2 + (y_j^s)^2} \tag{4-28}$$

对于链码 $\boldsymbol{L}_{j+s}^s$，同样有

$$\vec{\boldsymbol{L}}_{j+s}^s = x_{j+s}^s\vec{\boldsymbol{i}} + y_{j+s}^s\vec{\boldsymbol{j}} \tag{4-29}$$

$$|\vec{\boldsymbol{L}}_{j+s}^s| = \sqrt{(x_{j+s}^s)^2 + (y_{j+s}^s)^2} \tag{4-30}$$

显然，边界在点 j 处的曲率可由其两侧的局部链码向量 $\vec{L}_j^s$ 和 $\vec{L}_{j+s}^s$ 的夹角 θ_j^s 来近似计算，由于

$$\vec{\boldsymbol{L}}_j^s \cdot \vec{\boldsymbol{L}}_{j+s}^s = |\vec{\boldsymbol{L}}_j^s| \cdot |\vec{\boldsymbol{L}}_{j+s}^s| \cdot \cos\theta_j^s \tag{4-31}$$

可以推得

$$\theta_j^s = \cos^{-1}\left(\frac{x_j^s x_{j+s}^s + y_j^s y_{j+s}^s}{\sqrt{(x_j^s)^2 + (y_j^s)^2} \times \sqrt{(x_{j+s}^s)^2 + (y_{j+s}^s)^2}}\right) \quad (4\text{-}32)$$

可以求得所有边界点的曲率 θ_j^s，在整条链上的局部极大值位置就是角点。

由于提取的角点在轮廓的参照下，信息最为丰富，能构造出针对不同应用范围的特征向量。但该方法对前期的图像分割有很大的依赖性，而图像分割本身运算比较复杂，分割过程中出现的任何错误都有可能影响角点的检测。不过，在图像分割效果良好的情况下，这类方法简单实用，且有较高的检测精度和稳定性。

4.4.2 SUSAN 算法的自适应阈值改进

Smith 和 Brady[171] 提出的 SUSAN 算法是一种应用广泛的基于图像灰度变化的方法，随后出现的 MIC 算法[172] 等都是它的思想的改进和发展。该算法使用一个可调节大小的圆形模板，模板内的每一像素点灰度值与中心像素点灰度值比较，灰度值与中心像素点相近的点组成的区域，称为 USAN（Univalue Segment Assimilating Nucleus）区域。SUSAN 算法就是根据各个待考察点的 USAN 区域面积来判断当前点是区域内部点、边界点还是角点。

1. 基于 SUSAN 算法的角点检测

如图 4-7 所示，a 点模板处于背景中，整个模板都属于 USAN 区域；b 点有超过一半的像素点属于 USAN 区域；c 点模板内有一半像素点属于 USAN 区域；d 点有少于一半的像素点属于 USAN 区域。可见，如果待考察的像素点是角点，USAN 区域的面积最小。

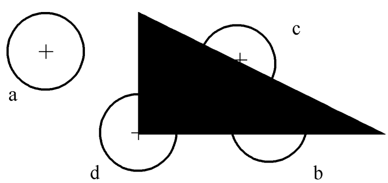


图 4-7 USAN 区域图解

图 4-8 所示为 SUSAN 算法的三种近似圆形模板，在实际应用中，37 邻域的 7×7 模板最为常用。

SUSAN 算法的数学描述为：使用近似圆形的模板（窗口）在图像上滑动，在每一个位置考察当前像素点的 USAN 区域面积。具体方法是比较窗口内的每一个点与中心点的灰度值差异：

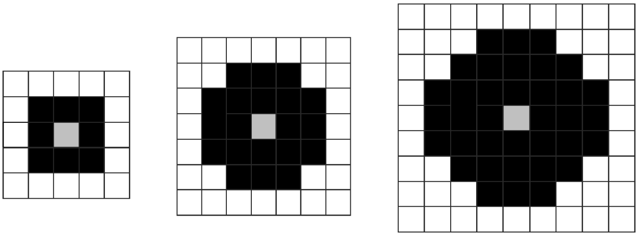


图 4-8　SUSAN 算法的模板

$$c(\vec{r},\vec{r_0})=\begin{cases}1 & \text{if}\,|I(\vec{r})-I(\vec{r_0})|\leqslant t\\0 & \text{if}\,|I(\vec{r})-I(\vec{r_0})|>t\end{cases}\tag{4-33}$$

为了得到更好的稳定性和有效性，可以用下面的判别函数：

$$c(\vec{r},\ \vec{r_0})=e^{-\left(\frac{I(\vec{r})-I(\vec{r_0})}{t}\right)^6}\tag{4-34}$$

式中，$\vec{r_0}$是当前像素点（中心点）的位置，$\vec{r}$ 是圆形窗口内其他任意一点的位置，$I(\vec{r})$ 表示 $\vec{r}$ 点的图像灰度值。t 则是预设的灰度差阈值，理论和实践都证明，一般指数为 6 时 t 取 25 效果最好。

计算以 $\vec{r_0}$为中心像素点的模板内 USAN 区域大小的公式表示如下：

$$n(\vec{r_0})=\sum_{\vec{r}}c(\vec{r},\ \vec{r_0})\tag{4-35}$$

接着，将 $n(\vec{r_0})$与预先给定的几何阈值 g 进行比较，可以得到图像的初始角点响应：

$$R(\vec{r_0})=\begin{cases}g-n(\vec{r_0}), & n(\vec{r_0})<g\\0, & n(\vec{r_0})\geqslant g\end{cases}\tag{4-36}$$

$R(\vec{r_0})$为反应函数，经过局部非极大值抑制 NMS（Non-maximum Suppression）之后确立为角点。因为在角点的一个邻域内往往不止一个点的 $R(\vec{r_0})$值大于零，只有 $R(\vec{r_0})$值最大的点才被确立为角点。

SUSAN 算法的优点是在角点检测时不需计算梯度，不需插值且不依赖于前期图像分割的结果，直接对像素的邻域灰度值比较即可检测出角点，速度比较快，有一定的抗噪声干扰能力。但是采用预设的固定阈值限制了该算法的适用范围，需要对其做相应的改进，使得它可以根据具体情况自适应地调整阈值。

2. 灰度阈值的自适应计算

在 SUSAN 算法中，几何阈值 g 和灰度阈值 t 的作用比较重要。几何阈值 g 决定了提取的角点的尖锐程度，g 越小提取的角点越尖锐。在用 SUSAN 算法进

行边缘提取的时候通常取 $g=3/4n_{max}$，进行角点提取的时候，则取 $g=1/2n_{max}$。

一般而言，对于 g 不需要通过调整就能取得较好的效果。灰度差阈值 t 决定了 SUSAN 算子所能检测到的最小的对比度以及去除噪声点的能力。t 越小，检测到的角点就越少，有可能漏检。t 越大，所能检测到的角点就越多，但有可能误检。因此，如果对于灰度细节比较丰富的图像使用统一的灰度差阈值 t，检测效果会不好。所以，需要有针对性地给出一种对 t 值的自适应的提取方法。

对于每个像素点的 SUSAN 模板，通过计算模板内每个像素点与中心点的灰度差得到该模板的灰度差直方图，然后根据灰度差直方图通过迭代法确定该模板的阈值 t，使得对于不同的对比度的图像都能够自适应的计算出每个模板内适合的 t 值。

首先计算模板中每点与中心点的灰度差阈值，然后取灰度差值的均值为迭代初始值 t_0，如下式：

$$t_0=\frac{1}{n}\sum I(r)-I(r_0) \tag{4-37}$$

然后根据迭代初值将灰度差直方图分为两部分，进行迭代计算：

$$t_{i+1}=\frac{1}{2}\left[\frac{\sum_{m=0}^{t_i} m\times h(m)}{\sum_{m=0}^{t_i} h(m)}+\frac{\sum_{m=t_i+1}^{C_{max}} m\times h(m)}{\sum_{m=t_i+1}^{C_{max}} h(m)}\right] \tag{4-38}$$

式中 m 为模板中像素点和中心像素点的灰度差值，$h(m)$为模板中具有该灰度差值的点的数量，C_{max}为灰度差值的最大值，迭代终止的条件是 $|t_{i+1}-t_i|=0$。

因为每个模板的 t 是根据模板内的灰度差值确定的，因此能够很好地检测到不同灰度对比度下的灰度变化，使得 USAN 区域的判断更加准确。

4.5　实验结果与分析

1. 实验环境

（1）硬件环境

普通 DELL 台式计算机一台，基本配置为 P（R）D/3.4GHz/1.00G/160G/19in[㊀]。

（2）软件环境

WindowsXP 操作系统，Visual Studio C++6.0 开发平台，OpenCV 函数库。

㊀ 1in=25.4mm

2. 实验数据来源

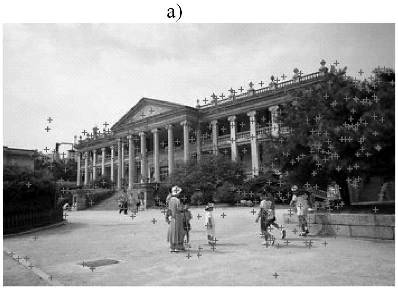
Caltech101 图像库共有 101 类目标，每类目标有 40 ~ 800 幅图像，图像大小为 300 × 200 像素；普林斯顿大学三维模型库（Princeton Shape Benchmark）[㊀]中的模型可以投影为 2D 图像，模拟相应物体分割后的二维灰度图像。我们从普林斯顿模型库中挑选出一个飞机模型 F16、一个坦克模型 T60、一辆汽车模型用来进行与角点检测相关的实验。

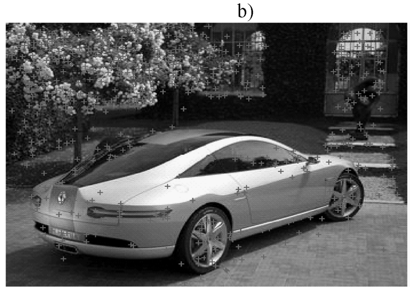
实验 1：DoG 特征点检测与 SIFT 特征描述子表示

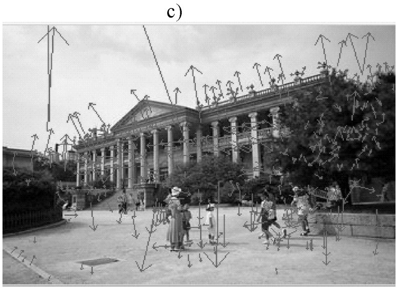
如图 4-9c 和 4. 9d 所示，DoG 检测算子是稀疏选取法的典型代表，其检测出


a)


b)


c)


d)


e)

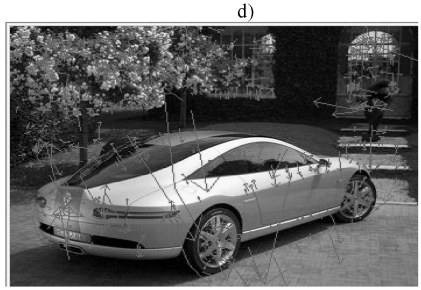

f)

图 4-9　图像局部特征的提取与表示

a）建筑物与人　b）汽车与树　c）建筑物与人的特征点　d）汽车与树的特征点　e）建筑物与人的特征描述　f）汽车与树的特征描述

㊀ http：//shape. cs. princeton. edu/benchmark。

的特征区域数量一般在200～3000个，其主要优点是简洁、紧致，图像的特征点远远少于图像的像素，使得后面的识别过程能大大加速。但很多特征区域检测算法往往和图像的特性相关，应用到通用目标识别时，可能会有一定的局限。图4-9e和4-9f为SIFT描述子的向量表示方式，箭头的起点代表该特征点的位置，箭头的长度代表该特征点所处的尺度，箭头的方向代表该特征点的主方向。

实验2：直线投影和SUSAN角点检测

如图4-10所示，直线投影法相对来说简单实用，具有较高的检测精度和稳定性，由于把角点定义在目标的轮廓线上，必须先分割图像并进行二值化，这样一来对前期的图像分割有很大的依赖性，而图像分割本身运算比较复杂，分割过程中出现的任何错误都有可能影响到检测结果。SUSAN算法则不需计算梯度，不需插值且不依赖于前期图像分割的结果，直接对像素的邻域灰度值比较即可检测出角点，速度比较快，通过自适应阈值的改进之后，该算法的抗噪声干扰能力有所加强，角点检测效果比较理想。

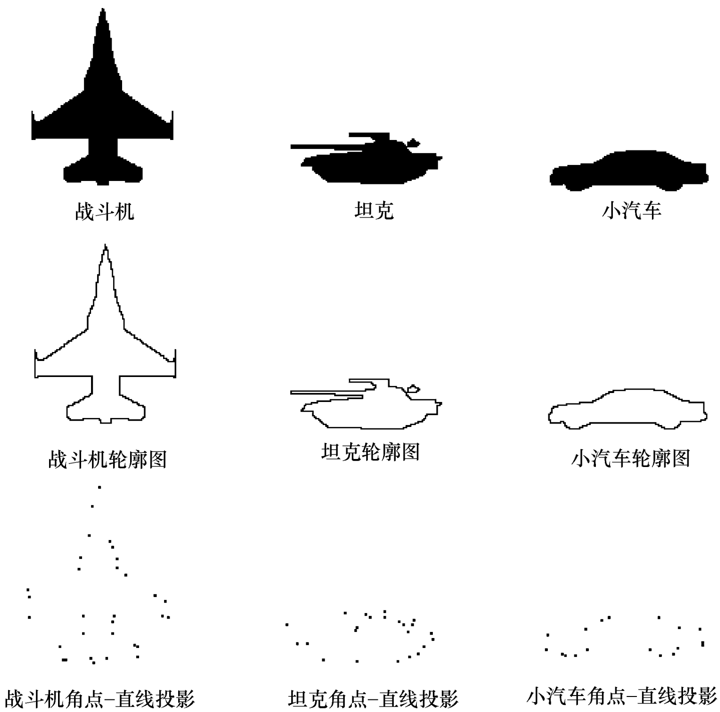
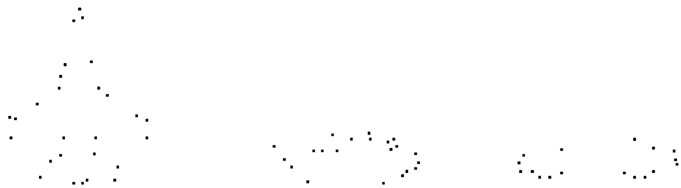


图4-10　二值图像的角点检测效果



战斗机角点-SUSAN　　坦克角点-SUSAN　　小汽车角点- SUSAN

图 4-10　二值图像的角点检测效果（续）

4.6　本章小结

本章将局部特征的提取作为目标识别的基础和关键进行了深入研究，根据本书的实验需要，在 DoG 特征点检测的基础上结合 SIFT 和 GLOH 描述子完成了对复杂图像的局部特征提取与描述；在狭义特征点——角点的检测技术研究中，针对 SUSAN 算子固定阈值的问题，提出了自适应阈值的改进方法，使得算法的应用更加灵活，检测结果更加稳定、有效。

最近几年国内外学者提出了许多局部特征提取方法，但现有的局部特征都有各自的局限性。随着在图像目标分类、图像目标检测等领域的深入应用，如何选择合适的局部特征来完成具体识别任务成为了算法的关键技术。而局部特征和整体特征各自都有优势和缺点，如果能将多种局部特征与整体特征结合起来，在目标识别领域将会有更好的发展前景和实用价值。

第 5 章　基于局部特征的目标匹配

在科学领域，我们探索真理的方法是：根据事实来设计实验，修改它们，然后再进行更多的实验。

5.1　引言

场景图像与照明、摄像机参数、摄像机位置等因素有关，因此，要从一幅图像中对目标进行匹配识别，特别是从复杂背景多物体的图像中识别特定目标，必须考虑这些因素：场景的不变性，场景的复杂度取决于获取图像时的条件（照明、背景、摄像机参数和观察点）是否同模型建立条件相似，场景的条件显著地影响同一物体的图像；图像模型空间，二维图像是三维物体在二维空间的映射，加之物体运动时的情况更为复杂；模型库中物体的数目，用于物体识别的特征选择计算量随着物体数量的增加而迅速增加；图像中的遮挡问题，遮挡导致原先特征点消失，新特征点的产生，因此在假设验证阶段就应该考虑遮挡问题。

局部特征的提出使得目标匹配可以从整体匹配的形式转变为局部匹配的形式，从而为遮挡目标的识别和不同姿态的同一目标的识别开辟了一条有效的途径。近些年来，随着基于局部特征的目标匹配方法的不断发展和

改进，其广泛应用于图像拼接和图像检索领域[173-176]，并取了阶段性的成果。本章在对国内外相关领域的众多研究成果进行深入探讨之后，针对局部特征匹配在目标图像拼接和图像检索中应用的不足，提出了基于多分辨率技术和局部特征的航拍图像拼接方法，以及基于原型匹配的图像检索方法。

5.2 结合 NNDR 与霍夫变换的匹配方法

在建立两幅图像之间局部特征的匹配关系时，可以参照 Marr 等人[177]提出的匹配应该满足唯一性、相似性、连续性三个基本约束条件，即物体表面任意一点到观察点的距离是唯一的，因此其视差是唯一的，给定一幅图像中的一点，其在另一幅图像中对应的匹配点最多只有一个；对应的特征应有相同的属性，在某种度量下，同一物理特征在两幅图像中具有相似的描述符；与观察点的距离相比，物体表面因凹凸不平引起的深度变化是缓慢的，因而视差变化是缓慢的，或者说视差具有连续性。

5.2.1 基于 NNDR 的匹配策略

目前常用的目标匹配策略有两种：一种是距离阈值法（Threshold-based Matching），即待匹配目标与模型之间的距离小于某个阈值，则认为匹配上了，该方法非常简单，但是阈值的确定非常困难，而且目标很容易匹配上多个模型，从而产生大量的误匹配；另一种是最小距离法（Minimum Distance），即目标只匹配与其距离最近的模型，实际应用中一般还需要满足距离小于某个阈值的条件，该方法只有一个最佳的匹配结果，相对于距离阈值法来说，正确率要高。

由于图像的内容千差万别，加上场景中的运动物体、不重叠内容以及图像质量等因素的存在，一幅图像中的局部特征并不一定能够在另一幅图像中找到相似的特征，这就需要采取措施剔除那些产生干扰的噪声点，通常把这样的点称为“外点”。许多图像的背景比较相似并不具有区分性，如天空、旷野之类，它们的局部特征之间的距离要小于有用的特征之间的距离，但是它们并不能描述图像的主要内容，所以设置一个全局性的距离阈值来决定局部特征匹配与否显然是不合适的。

对 SIFT 特征的研究[146]表明，可以通过比较最近邻（First Nearest Neighbor）

特征和次近邻（Second Nearest Neighbor）特征的距离可以有效地甄别局部特征是否正确匹配。这就是最邻近距离比值法（Nearest Neighbor Distance Ratio，NNDR），其表述如下，如果待匹配特征为 D_A，其最邻近特征为 D_B，次邻近特征为 D_C，那么判断该特征匹配的条件为：

$$\frac{\|D_A - D_B\|}{\|D_A - D_C\|} < t \tag{5-1}$$

该方法理论来源是，如果一个特征在一幅图像中与两个特征的距离都很相近，那么该特征的区分度较低，也违背了 Marr 提出的“匹配应该满足唯一性”的原则，会对图像相似度的判断产生干扰。如图 5-1 所示，进行 SIFT 特征匹配的实验结果也证实了这一点，当剔除与最近邻点和次近邻点距离比值大于 0.8 的特征对时，排除了 90% 的干扰而仅仅误删了 5% 的正确特征对。

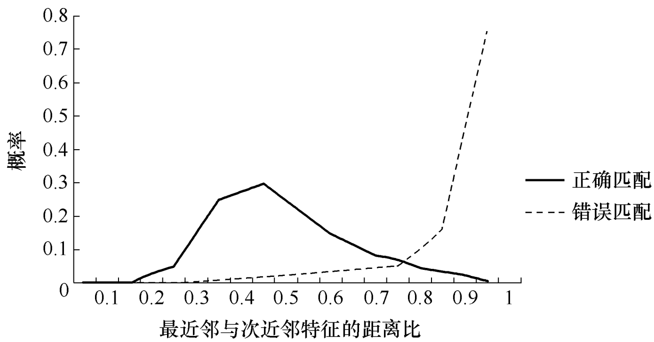


图 5-1　特征点匹配的概率分布

该实验从图像数据库中选取的待匹配图像，共提取了 40000 个 SIFT 特征，对这些待匹配的图像进行了随机数值的尺度变化和平面旋转，并进行了深度小于 30°的视角变化处理，同时也加入 2% 的高斯噪声。

5.2.2　邻近特征点的搜索算法

用穷举法搜寻最邻近点以及次邻近点，可以得到最精确的结果。但是由于本书所用的特征空间一般都高达 128 维以上，加之复杂图像的局部特征数量比较多，搜索算法的效率显然成为了整个系统的一个瓶颈。

1. K-D 树搜索策略

标准 K-D 树是 Friedman 等人[178]提出的一种高维二叉树，K 表示空间的维数，在其上可实现对给定特征点的快速最近邻查找。若某 K-D 树的结点数目为

N，则在它上面的最邻近节点的平均计算复杂度为 $O(\lg N)$。其后又相继提出了K-D 树的 $1+\varepsilon$ 近似最近邻搜索算法（ANNS），其主要思想是在搜索时只查询那些与给定特征点的距离小于当前最近距离 $1/(1+\varepsilon)$倍的点，此时搜索完成时返回的点未必是真实的最近邻点（除非 $\varepsilon=0$），但是即使当 ε 取的较大（如 $\varepsilon=3$ 时），所返回的点仍然有 50% 的机会是真实的最近邻点，而且在平均意义上它们到目标点的距离只是真实最近邻点到目标点的距离的 11.5 倍，取得的加速比却可以达到 50 倍以上。

在数据维数较低的时候，K-D 树搜索方法比较有效。在更高维的数据空间中将会有更多的分类结果接近目标真实数据，此时使用 K-D 树进行搜索的话，效率将会急剧下降。本书所用的特征空间一般都高达 128 维以上，为此本书使用了 Beis 和 Lowe 提出的 Best-Bin-First（BBF）算法[146]，它对常规的 K-D 树搜索方法进行改进，从而实现较快的匹配点搜索。

2. 基于 BBF 算法的搜索策略

在高维数据搜索空间中，K-D 树搜索的结果仅仅只有很少的一部分满足邻近原则，为了加快搜索速度，可以通过减少搜索节点来缩小搜索范围。这需要使用一个基于堆的优先级队列，将搜索空间的节点按照与待查询节点的距离来进行排序。当搜索到的节点符合设定的约束条件，则记录到优先级队列中去，从而获取下一个候选节点的信息（包括该节点在当前树的位置和到待查询节点的距离）。当一个最邻近点被搜索到后，则从队列的队首删除一项，然后继续搜索包含最近邻节点的其他分支。

如图 5-2 所示，对于特征点数量达到 10000、维度为 5 ~ 25 的数据检索中，按照 BBF 算法改进的搜索策略很大程度上提高了检索效率，而标准的 K-D 树搜索策略在数据维度达到 10 后其效率便明显下降了。

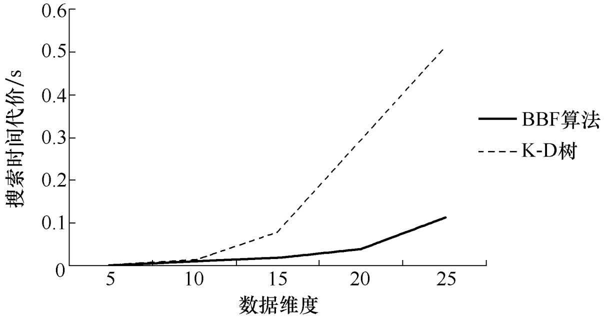


图 5-2　BBF 算法与 K-D 树的搜索时间代价

本书设定的约束条件是检查前 200 个最邻近候选节点，该算法在搜索速度提高了 2 个数量级的同时，平均只丢失 5% 的特征对，这对于一般的图像检索来说是可以容忍的。当距离非常相近的特征点需要进一步甄别的时候，BBF 算法的搜索效率会受到制约，但是本书剔除了与最近邻点和次近邻点距离比值大于 0. 8 的特征对，这就基本上避免了这一困境。

5. 2. 3　基于霍夫变换的目标检测

一幅图像往往可以提取出超过 2000 个局部特征，而这些局部特征很可能来自场景中的多个物体或背景。如何从这些特征中寻找到只属于待识别目标的局部特征子集，这是进行目标匹配识别所必须解决的问题。霍夫变换（Hough Transform）为此提供了一条高效的途径。

基本的霍夫变换最初是用来进行直线检测的，而广义霍夫变换则可以在所需检测的曲线或目标轮廓没有或不易用解析表达式时，利用表格来建立曲线或轮廓点与参考点间的关系，进而检测出目标[67]。霍夫变换的基本思想是将原图像变换到参数空间，用大多数边界点满足某种参数形式来描述图像中的线，通过设置累加器进行累积，求得峰值对应的点所需要的信息。霍夫变换以其对局部缺损的不敏感，对随机噪声具有鲁棒性以及适于并行处理等优良特性，备受图像处理、模式识别和计算机视觉领域学者的青睐。霍夫变换的突出优点就是可以将图像中较为困难的全局检测问题转换为参数空间中相对容易解决的局部峰值检测问题。

霍夫变换利用点线对偶性原理进行坐标变换，原理如图 5-3 所示，在直角坐标系下，利用公式（5-2）表示过点（x，y）的直线 L_0 的方程：

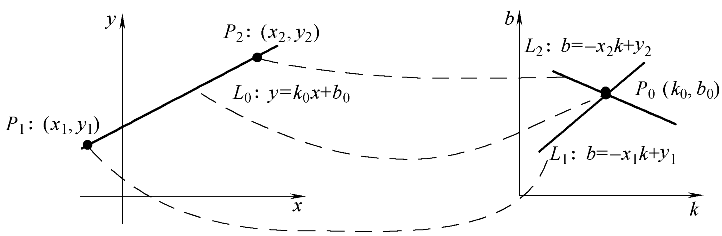


图 5-3　直线检测中的霍夫变换

$$y = k_0 x + b_0 \tag{5-2}$$

式中，k_0 为斜率；b_0 为截距。将其变换为参数空间中过点（k_0，b_0）的直线方程：

$$b_0 = -xk_0 + y \tag{5-3}$$

可以看出，直线 L_0 上的两个点（x_1，y_1）和（x_2，y_2），在参数空间中表示为两条直线不同的直线 L_1 和 L_2，而它们在参数空间中相交于（k_0，b_0）点。也就是说，原图像空间中同一条直线上的不同点在参数空间中被变换为一组相交于同一点的直线。

使用公式（5-2）表示一条直线带来的一个问题是，当直线接近垂直时，直线的斜率接近无限大。解决这一难点的一种方法是使用极坐标方程来表示直线：

$$\rho = x\cos\theta + y\sin\theta \tag{5-4}$$

其中 ρ 为原点到直线的距离（即原点到直线的垂直线的长度），θ 确定了直线的方向（即原点到直线的垂直线与 x 轴方向的夹角）。如果对位于同一直线上的 n 个点进行霍夫变换，则原图像空间中的这 n 个点在参数空间中对应得到 n 条正弦曲线，并且这些曲线相交于同一点了，若能确定参数空间中的 P_0 点（局部最大值），也就实现了直线的检测。

本书对目标姿态建立一个参数空间，将目标的 2D 坐标、尺度、方向参数等坐标轴按照一定的步长划分为若干等份；然后将所有匹配的特征点向这个参数空间投票；对参数空间每个点的投票累加值进行分析，累加值大的点所对应的目标姿态有更高的概率出现在图像中。在实际应用中，对于参数空间坐标轴步长，一般 2D 坐标为训练集中目标最大尺寸的 0.25 倍，尺度因子为 2，方向参数为 30°。累加值和预设的阈值进行比较，当大于阈值时，则判定该点所对应的目标姿态存在于图像中。

在图 5-4 中，左图所示的两个目标——玩具火车和玩具青蛙，由于其他物体（包和箱子等）的存在，在中间的图像里都产生了局部遮挡，而采用上述的目标匹配方法都可以将这些目标识别出来，识别效果图如右图所示。大的矩形框中是识别出的目标，小的矩形框代表识别所用到的局部特征。



图 5-4　局部遮挡目标检测（来源： Lowe, 2004 年）

5.3　基于局部特征和多分辨率技术的图像拼接

航拍图像拼接技术是当前机器视觉领域的一个研究热点，已经被广泛应用于地理信息系统、地质灾害监测、城市规划和战场态势评估等许多方面。其主要内容和一般的图像拼接一样，就是将一组互相有重叠部分的图像序列进行空间配准，拼成一幅包含各图像序列信息的宽视角、完整的新图像，以满足现实要求[179]。但是，由于是在飞行器上对地面场景的俯视拍摄，所以又有其自身的特点和难点，比如飞行器姿态变化导致的航拍视角改变、飞行器升降造成的图像分辨率不同、天气状况对图像质量的影响等。

图像拼接方法通常可以分为两类[173,180,181]，一类是将场景投影到柱面坐标下进行拼接，这类方法模型简单且计算速度快，但是要求相机只能围绕光心做水平旋转运动，还需要获取拍摄每幅图像的焦距。该方法比较适合于全景图像拼接。另一类方法则是以仿射变换模型为理论基础，广泛应用于航拍图像拼接，一般首先需要根据飞行器和相机的参数计算图像的位置坐标并排列图像，然后检测相邻图像重叠区域内的对应点以求得图像间的变换关系。而正如前面所述，航拍图像的特殊性使得图像的位置坐标不准确，有时还需要从航拍视频中抽取图像进行拼接，这就需要一种更为稳健高效的拼接方案。

5.3.1　图像拼接技术的研究现状

早在 1992 年，英国剑桥大学的 Lisa Gottesfeld Brown 在文献中就总结了图像配准的主要理论及图像拼接技术在各个领域的应用，当时他的讨论主要还是着眼于医学图像处理、遥感图像处理等传统应用领域。时隔 20 年，图像拼接技术有了飞跃发展，目前在大面积场景观测、虚拟现实、视频压缩、视频检索以及高分辨率图像的获取方面也有了广泛应用。

1. 大面积场景观测及视频监控系统

图像拼接技术可以用于场景观测，通过将卫星图片或航空照片或者水下摄像图片拼接成大范围的场景图片来实现对某一地区某一场景的整体勘察观测，比如高大建筑物高分辨率全景图像的获取、水下考古、海底探测以及遥感观测等。视频图像序列构造全景视图技术还可以用于现场操作员和指挥专家之间的远程协作系统和远程遥控系统，现场操作员通过头盔摄像机将现场拍摄的视频图像通过无线通信的方式传递给在远程的指挥专家，远程指挥专家在收到现场

拍摄的视频图像后构建出现场的全景图像，然后根据现场情况提出建议并通知现场操作员进行相关的操作。

2. 虚拟现实场景的构建

虚拟现实技术是利用计算机构建一个逼真的虚拟环境，即以仿真的方式给人们创造一个反映实体对象变化及其相互作用的三维世界，使得人们能够通过使用专用设备，就能像在自然环境中一样对虚拟环境中的实体进行观察与控制。在20世纪90年代，由于传统的基于图形绘制（GBR）的虚拟现实技术存在着明显的缺点，无法完全适应实际需要，人们提出一种基于图像绘制（IBR）的虚拟现实技术，通过许多相关的静止的图像进行连续的插值而实现场景的交互式浏览，这样大大降低了数据量，从而方便了图像数据的传输和保存。虚拟现实技术所需要的图像依赖于图像拼接技术，所以图像拼接技术有重要的研究价值。

3. 视频压缩

图像拼接技术的另外一个重要应用是视频压缩。目前 MPEG-4 编码标准针对视频中背景对象的特点提出了 Sprite 编码方式。利用图像拼接技术将整个视频图像序列的背景内容拼接成一幅大的完整的背景全景图像，该背景在每一帧中出现过的像素点，在这幅大的背景全景图中都能找到对应的点，这样的图像就叫做 Sprite 图像。由于 Sprite 图像自身是不变的，因此只需传输一次，然后根据摄像机的运动参数在接收端重建背景，这样可以大大减少传输的数据量。这种编码方式可以很大程度上提高视频压缩效率。

4. 视频检索

视频流帧间存在大量冗余信息，利用图像拼接技术去除冗余，将分散在各个视频帧中的信息集中起来表示成整体的场景，这种紧密重组提供了对内容的非线性浏览和高效的索引，可以有效地对感兴趣的信息进行直接快速存取、编辑注释等操作。

从具体算法角度来讲，国际上在1996年由 Richard Szeliski 提出了基于运动的全景图像拼接，该算法是图像拼接领域的一个里程碑式算法。它是采用了 Levenberg-Marquardt 最优化算法使得两幅图像的亮度差最小，进而求出图像间的变换关系，此方法效果比较理想，还可以处理平移、旋转、仿射等多种图像变换。而 Richard Szeliski 也成为了图像拼接领域的奠基人，这套理论已经成为了一个经典理论体系，现在许多人依然在这套理论基础上做进一步研究。2000年，Shmuel Peleg，Benny Rousso 等人做了进一步的改进，提出了一种自适应的图像

拼接算法，它是依据摄像机的不同运动方式，自动选择合适的拼接模型。这一研究成果推动了图像拼接技术的发展，由此，拼接的自适应性成为图像拼接领域研究的热点。在 2003 年 ICCV 大会上，M. Brown 发表了一篇名为《Recognising Panoramas》的文章，文中使用了基于尺度不变特征的匹配算法进行图像拼接，并采用多分辨率的思想进行图像融合，将低频信息与高频信息采用不同的方式进行融合，既保证了细节信息，也保证了背景信息，该算法的自适应性好，并且效果理想。因此 M. Brown 提出的理论大大地推动了图像拼接技术的发展，也将全景图拼接技术研究推向高潮。

国内关于图像拼接技术的研究也发展较快。1997 年，浙江大学 CAD&CG 国家重点实验室研究并提出一种自动拼接算法，该算法是基于模板匹配的思想进行搜索，确定最佳匹配方式。1998 年，Paul Bao 运用小波变换的优良性质提出一种图像拼接算法，该算法结果精度高，拼接效果好，但是小波变换同傅里叶变换一样存在效率低的缺点，需要进一步改进。2001 年，清华大学的研究人员提出了一种新的图像拼接算法，研究算法效率与精确度的关系，将摄像机固定在特殊的三脚架上，使其绕垂直轴旋转拍摄，最终取得了不错的拼接效果。同年，华中科技大学的研究人员通过研究图像变换关系模型，提出了基于特征点的改进拼接算法，它是首先运用相关法提取特征点，再计算变换模型生成全景图的算法。2002 年，杜威等人对动态全景图做了相应研究，提出了一种能够处理动态场景的全景图表示方法，把视频和全景图结合起来，生成动态全景图。在国内比较优秀的拼接算法是在 2004 年由赵向阳、杜立民提出的一种基于特征点匹配的拼接算法，它首次将角点匹配与变换参数鲁棒估计引入图像拼接，虽然说大部分都是国外经典算法，但是该论文的主要贡献是将这些算法有机地组合起来，并取得理想效果。在此基础上，2008 年，马丽涛等提出一种基于条件数的配准算法，其主要思想是：在角点特征的基础上，研究分析噪声对图像之间的变换关系的影响程度，然后筛选出具有稳定性的角点，提高了匹配的准确度。

5.3.2　多分辨率下的图像配准

图像配准也称图像对齐，是对从不同传感器或不同时间或不同角度所获取的两幅或多幅图像进行最佳匹配的处理过程。而图像配准的本质是寻找一种图像对之间的变换关系，在这种变换关系下，两幅图像之间可以建立像素点之间的对应关系。经过多年发展，人们提出了许多种图像配准的方法[174,182,183]，大体

可以分为三类。

1. 基于频域的方法

基于频域的方法，即相位相关法。它是利用傅里叶变换将两幅待配准的图像变换到频域，然后利用它们的互功率谱直接计算图像的变换关系，从而完成配准。其优点是算法简单，效果理想，图像存在的平移、旋转、仿射等变换关系会在傅里叶变换域上有相应的体现，所以该类方法具有一定的鲁棒性。拼接的前提条件是待拼接图像之间重叠区域比例大，一般要求超过 50%，这使得其在实际运用受到较大限制。

2. 基于区域的方法

基于区域的方法，即灰度相关法。它是计算图像之间重叠区域对应灰度的统计信息，然后根据特定的相似度量为配准准则。该类方法实现简单，但是应用范围非常狭窄，不能用于非线性变换，而且运算量大。

3. 基于特征的方法

提取图像的局部特征信息，运用特定的相似度量实现配准。由于图像特征种类非常多，有特征点、边缘、轮廓、闭合区域、统计特征等，相对于其他方法，基于特征的方法运算速度较快，能够容忍较大的图像差异，获得的配准结果比较稳定，已经成为当前主流的图像配准方法。

当前已有的基于特征的图像配准方法普遍存在一个问题：它们提取的特征稳定性较差，通常不具备对仿射或透视投影变换的不变性，难以适用于成像情况相对复杂的航拍图像。近年来，在工程应用中发现，局部特征不仅对图像尺度、平移、旋转变换具有不变性，而且对光照变化以及复杂的投影变换也具有部分不变性，比较适合用于航拍图像序列的处理，在图像场景较大、天气和飞行器姿态的影响普遍存在的情况下，可以实现准确、稳健的航拍图像配准。

许多国内外文献，如参考文献［143，144，146］都曾指出，在复杂内容的图像中提取的特征点非常多，过多的特征点不仅会加重计算负担，影响效率，而且会对特征匹配造成干扰，不利于航拍图像序列的准实时拼接。本书的拼接方法只需利用少量（3 个以上）特征点即可完成图像配准，这对特征点的提取质量提出了较高要求，而多分辨率分析就为解决这个问题提供了一条有效的途径。

当观察图像时，通常看到的是相连接的纹理与灰度级相似的区域，它们相结合形成物体。如果物体的尺寸很小或对比度不高，通常采用较高的分辨率观察；如果物体尺寸很大或对比度很强，则只需较低的分辨率。如果物体的尺寸

有大有小，或对比度有强有弱的情况同时存在，以若干分辨率对它们进行研究将具有优势。这就是多分辨率处理的魅力所在，而且这样由粗糙到精细的分析策略在模式识别中可以发挥出很大的作用。

以多分辨率来解释图像的一种有效但概念简单的结构就是图像金字塔[8]。图像金字塔最初用于机器视觉和图像压缩，一幅图像的金字塔就是一系列以金字塔形排列的分辨率逐步降低的图像集合。如图 5-5 所示，金字塔的底部是待处理图像的高分辨率表示，顶部是低分辨率的近似。当向金字塔的上层移动时，尺寸和分辨率降低。因为基础级 J 的尺寸是 $2^J \times 2^J$ 或 $N \times N$（$J = \log_2 N$），所以中间级 j 的尺寸是 $2^j \times 2^j$。完整的金字塔由 $J+1$ 个分辨率级组成，由 $2^J \times 2^J$ 到 $2^0 \times 2^0$，但大部分金字塔只有 $P+1$ 级，其中 $j = J-p$，…，$J-2$，$J-1$，J 且 $1 \leqslant P \leqslant J$。也就是说，通常限制它们只使用 P 级来减少原始图像近似值的尺寸。

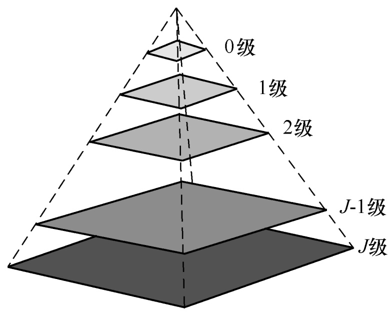


图 5-5　图像金字塔的结构

如图 5-6 所示，由于从机载摄影器材上获取的图像分辨率较高，本书通过建立图像金字塔来降低待匹配图像的分辨率，在低分辨率的图像序列上提取出更具代表性的特征点对，并计算出这些特征点在原始图像中的位置从而进行图像变换。

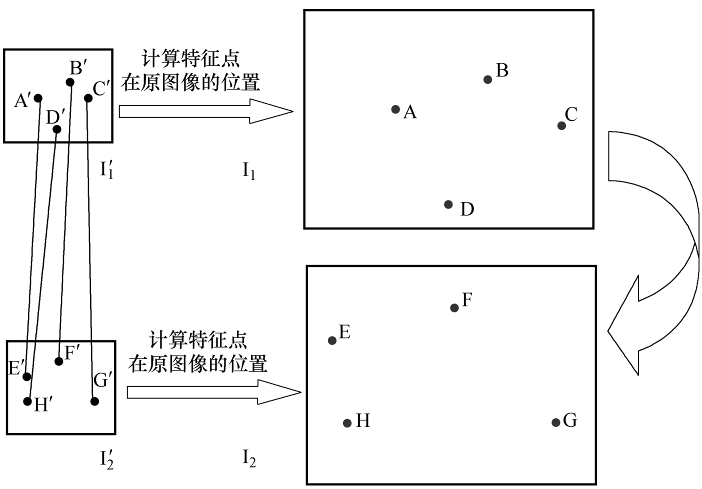


图 5-6　基于多分辨率技术的图像配准

为了实现图像序列的拼接，就必须确定有重叠的两幅相邻图像的空间对应关系，这也是图像拼接中关键的一步。为了确定图像间的对应关系，需要清楚相机进行图像采集时的运动方式，因为相机不同的运动方式会对场景成像产生不同效果，如图 5-7 所示。

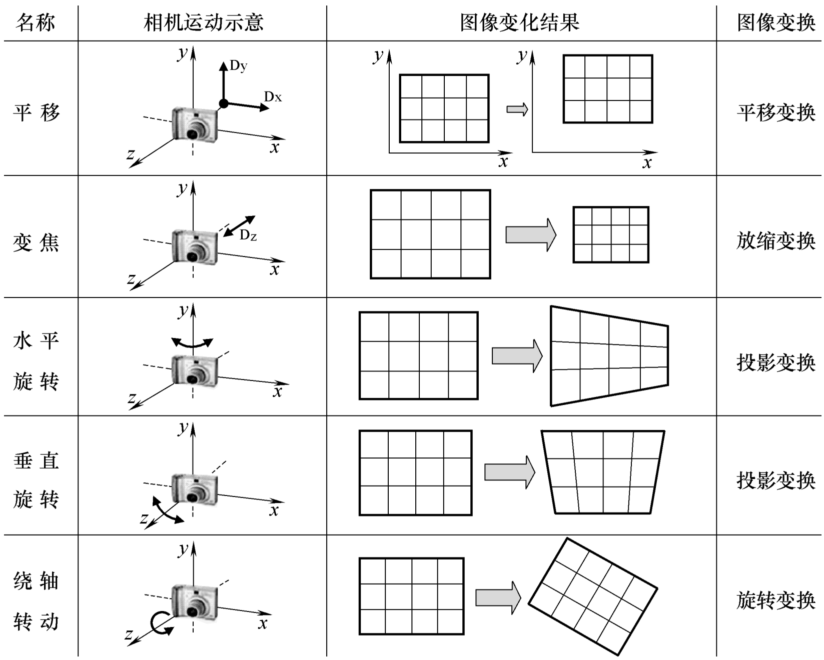


图 5-7　相机的运动方式与成像结果之间的关系

一旦确定了图像间的关系模型，则图像之间的配准问题就转化成确定该模型的参数问题。目前常用的关系模型有刚性变换（Rigid Transform）模型、仿射变换（Affine Transform）模型、投影变换（Projective Transform）模型以及非线性变换（Nonlinear Transformation）模型等。

1）刚性变换：如果一幅图像中的两点间的距离经变换到另一幅图像中后仍然保持不变，则这种变换称为刚性变换。刚性变换只局限于平移、旋转和反转（镜像），不会扭曲物体的原有形状，其变换矩阵具有 3 个自由度。

2）仿射变换：如果一幅图像上的直线经过变换后映射到另一幅图像上仍然为直线，并且保持平行关系，则这种变换称为仿射变换。仿射变换描述摄像机的平移、旋转、缩放运动。其变换矩阵具有 6 个自由度。

3）投影变换：如果一幅图像上的直线经过变换后映射到另一幅图像上仍然为直线，但平行关系基本不保持，则这种变换称为投影变换。投影变换具有更

一般的形式，可以描述摄像机的平移、水平扫动、垂直扫动、旋转、镜头缩放等运动，其变换矩阵具有 8 个自由度。它适用于景物平面相对于像平面有一定倾斜的情况，刚性变换模型和仿射变换模型可以看做是投影变换模型的特例。

4）非线性变换：非线性变换，也称为弯曲变换。经过非线性变换，一幅图像上的直线映射到另一幅图像上不一定是直线，可能是曲线。多项式变换是典型的非线性变换，如二次、三次函数及样条函数，有时也使用指数函数。

理论上讲，在图像变换的时候考虑的参数越多，得到的结果越精确。但在实际应用中，由于飞行器飞行轨道的起伏、地面物体高度的变化等因素，参数过多的变换矩阵反而起到的放大误差的效果，并且需要至少 7 个特征点对才可以进行配准。

通过对实际数据的研究，我们发现航空拍摄平台通常距离地面较远，可以将一定范围内的大地场景近似看成一个平面区域，这样一来就能够把一定长度的航拍图像序列变换到同一个成像平面完成图像配准。在各种图像变换模型中，虽然投影变换的描述能力更强，但依据奥卡姆剃刀（Occam's Razor）定律㊀，本书针对航拍图像的特点采用了仿射变换模型。该模型可以描述图像的旋转、平移和缩放等运动，利用 3 个以上特征点即可完成图像拼接，不仅极大简化了计算，拼接的最终效果也能够达到相应要求。

设成像平面上某一点 P_i 的坐标为 (x_i, y_i)，其三维齐次坐标为 $(x_i, y_i, 1)$。设一个观测点在两个相邻帧图像上所成的像点分别为 P_1 和 P_2，则这两点的齐次坐标之间满足如下关系：

$$P_2 = \boldsymbol{T}P_1 \tag{5-5}$$

式中，$\boldsymbol{T}$ 为 8 参数投影变换矩阵。实验证明，由于航拍图像序列中相邻两帧图像间视差较小，可以用式（5-6）给出的仿射变换矩阵来近似表达式（5-5）中的 $\boldsymbol{T}$，这样也有效地简化了计算。

$$\boldsymbol{T} = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ 0 & 0 & 1 \end{bmatrix} \tag{5-6}$$

根据线性方程相关理论，至少需要 3 个特征点对才能对这个 6 参数的仿射变换矩阵求解。但实际应用中，我们提取的特征点对数量通常大于 3，这时可以利

㊀ 该定律是 14 世纪逻辑学家、圣方济各会修士奥卡姆的威廉提出的，其原理的一种表述为“如无必要，勿增实体”。

用最小二乘法估计仿射变换矩阵 $\boldsymbol{T}$，相应的误差为

$$\mathrm{MSE}=\frac{\sum_{i=1}^{n}\|P_2-TP_1\|_2}{n} \tag{5-7}$$

判断图像配准的结果优劣的标准与其应用的领域有关系。比如，在军事制导领域，图像中目标定位的精确度与算法的速度是最重要的；在医学领域，获取的图像简单而正规，就可以采用比较简单的模板匹配；在卫星遥感方面，可以采用已知位置的标定物来定位配准。可见，图像配准本身就具有多样性和特殊性，在这几十年的技术发展过程中，还存在许多问题。图像配准的精度和效率上很难找到一个通用的平衡点，其针对性较强，自适应性不足，限制了图像拼接的实际应用范围。

5.3.3　渐入渐出的图像融合算法

由于进行航拍图像序列采集时拍摄条件的变化以及配准误差等因素的影响，叠加后的图像将不可避免地存在如光照变化、色彩差异、几何形变等诸多问题，从而在拼接结果中引入一些视觉上不连续的条带。如何消除这种拼接痕迹，使得图像过渡更加自然，这正是图像融合技术着力解决的难题。

现有的图像融合技术通常在像素级、特征级和决策级三个层次进行，如图 5-8 所示。其中，像素级图像融合是在基础层面上进行的信息融合，也是目前在实际中应用最广泛的图像融合方式，其思想是直接进行图像信息的综合而得到融合图像。图像拼接中的融合主要针对两幅图像重叠区域的平滑过渡，一般不需要进行高层次的数据融合，只是在像素级上进行处理就可以了。

目前常用的图像融合方法主要有直接平均融合、加权平均融合和多分辨率融合等。直接平均融合是将配准后图像之间的重叠区域对应像素点的灰度值直接进行叠加再求平均，相当于对图像进行了低通滤波。该方法简单但是通用性较差，最终图像中往往有较为明显的拼接痕迹，如果场景中存在运动目标还会产生“鬼影（ghost-like）”现象；多分辨率方法采用图像金字塔结构，将原始图像分解成不同频率上的一组图像，在每个分解的频率上，将图像重叠边界附近加权平均，最后把所有频率上的合成图像汇总成一幅。在每一个频率带内，加权函数的系数以及颜色融合区域的大小，是由两幅图像的图像特征在该频率带内的差异决定的，这样可以使得具有不同强度的图像平滑的过渡。虽然该方法拼接质量很高，但是计算过于复杂，不适宜大场景的准实时拼接；本书采用了加权平均方法中的渐入渐出融合算法，在保证航拍图像序列准实时拼接的同时，

实现了图像内容的平滑过渡。

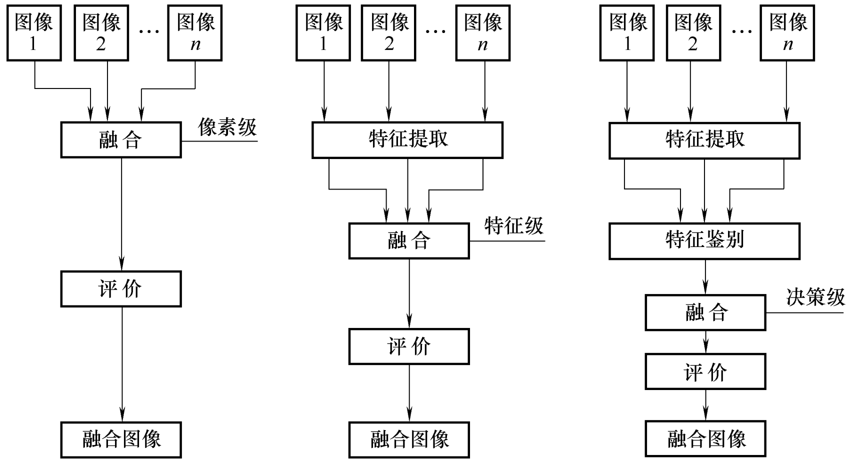


图 5-8　图像融合的层次

渐入渐出融合算法是由 Szeliski[184] 提出的，设 f_1 和 f_2 是两幅待拼接的图像，将它们按照线性加权的方法进行融合，融合后的图像像素 f 可表示为

$$f(x,y)=\begin{cases} f_1(x,\ y) & (x,\ y)\in f_1 \\ d_1f_1(x,\ y)+d_2f_2(x,\ y) & (x,\ y)\in(f_1\cap f_2) \\ f_2(x,\ y) & (x,\ y)\in f_2 \end{cases} \tag{5-8}$$

式中，d_1 和 d_2 表示与重叠区域的宽度有关的权重值，并且 $d_1+d_2=1$，$0\leqslant d_1$，$d_2\leqslant 1$。假设当前像素的横坐标为 x_i，重叠区域左右边界的横坐标分别为 x_l 和 x_r，那么在重叠区域中 d_1 由 1 渐变至 0，d_2 由 0 渐变至 1。

$$d_1=\frac{x_r-x_i}{x_r-x_l} \tag{5-9}$$

在对单程航拍的图像序列进行拼接时，采用上述方法，仅针对相邻两幅图像的 x 方向上作了平滑过渡，基本可以满足需求。如果图像序列是盘旋拍摄或者沿“几”字形路线拍摄的情况下，就需要进一步考虑到多幅图像、各个方向上的平滑[185]。本书在融合算法中为图像的每个像素分配权重，这个权重与像素到图像边缘的距离成正比：

$$C(x,\ y)=\frac{\sum_k w(d(x,\ y))I_k(x,\ y)}{\sum_k w(d(x,\ y))} \tag{5-10}$$

式中，w 是单调函数，一般取 $w(x)=x$，$I_k(x,\ y)$ 是第 k 幅图像在（x，y）点的

灰度值。$d(x, y)$的计算可以简单地取（x, y）点到图像四条边的最小距离。

5.4　基于局部特征和原型匹配的图像检索

现有的图像检索方式主要分为两种：基于文本的图像检索（Text-Based Image Retrieval，TBIR）和基于内容的图像检索（Content-Based Image Retrieval，CBIR）。前者自20世纪70年代发展至今取得了一定的成果[186,187,188]，但是三个突出的局限性使得它很难适应现实的要求：海量数据的标注耗时费力；主观性强，不同的理解导致对同一图像的标注差异很大；图像丰富的内容很难用少量文字描述清楚。

而基于内容的图像检索技术则通过图像的颜色、纹理、形状等视觉特征实现了“以图找图”的查询模式，其处理过程融合了图像分析、模式识别以及人机交互等多种技术，从20世纪90年代开始，逐渐成为了图像检索方向的研究热点[189,190,191]。随着在生产生活中的大量应用，基于内容的图像检索方法也显现出了一些不足，一方面是目前常用的图像特征大都是整体特征，如不变矩、纹理、欧拉向量、颜色直方图等，不能准确地表达场景信息和物体的本质属性；另一方面，由于图像理解技术的局限和用户界面的限制，检索系统给出的初始结果往往不能很好地满足用户的信息需求。

针对以上两点问题，本书对局部特征提取技术和相关反馈技术进行了深入的研究分析，提出了一种基于局部特征的图像检索方法。实验结果表明，该方法效果良好、性能稳定，有很大的发展潜力和广阔的应用前景。

5.4.1　CBIR的研究现状和发展趋势

国内外的研究机构已经投入大量人力物力开展了基于内容的图像检索方面的广泛研究，并且研制出了一些商业系统和实验系统。常见的基于内容的图像检索系统包括由IBM T. J. Watson研究中心开发的颇具影响力的QBIC系统、由哥伦比亚大学研究开发的VisualSEEK和WebSEEK系统、由美国Virage公司开发的Virage系统、由美国MIT媒体实验室开发的Photobook系统、由美国斯坦福大学研制的SIMPLIcity系统等，近年来国内也有一些大专院校研究开发了基于内容的图像检索系统，如浙江大学开发了基于图像颜色的检索系统PhotoNavigator，并将基于颜色的图像检索技术较为成功地应用于敦煌壁画数据库的研究和开发，复旦大学研制出iFind系统等。

QBIC 系统[一]是由 IBM Almaden 研究中心开发的第一个商品化的基于内容图像检索系统，它的系统框架、结构和技术对后来的图像检索系统有着深远的影响。QBIC 系统支持基于例子图像、手绘略图、选择的颜色、纹理等的查询，不仅支持图像检索，还支持视频、文本和语音多种形式的信息检索。QBIC 是少数几个考虑高维特征索引的系统。QBIC 系统使用的颜色特征是颜色直方图。纹理特征采用粗糙度、对比度和方向性描述。形状特征包括面积、圆形度、离心率、主轴方向和不变矩。颜色、纹理和形状均采用加权的欧式距离比较。

Virage[二]是由 Virage 公司开发的基于内容的图像搜索引擎。与 QBIC 相似，它支持基于颜色、颜色布局、纹理及结构的查询，但比 QBIC 更进一步的是它还支持上述四种特征的组合查询，用户可以根据自己的爱好调整这四种特征的权重。Virage 技术的核心是 Virage Engine 以及在图像对象层上的操作。Virage Engine 主要有图像分析、图像比较和图像管理三方面的功能。它将查询引擎作为一个插件，既可以应用到通用的图像查询中，也可对其进行扩展并应用到特定的领域。

Photobook[三]是 MIT 多媒体实验室开发的用于浏览和搜索图像的一套交互式工具。Photobook 包括三部分，形状提取部分、纹理提取部分及面部特征提取部分。它的人脸识别检索技术已被用于美国的警察机关。由于没有哪一种最好的特征能够单独地描述一幅图像，所以在 Photobook 的最新版本 FourEyes 中，Picard 等人提出了把用户加入到图像注释和检索过程中的思想。同时由于人的感知是主观的，他们又提出了把“模型集合”和人的因素相结合。实验结果表明，这种方法对于交互式图像注释来说非常有效。

VisualSEEK[四]是基于视觉特征的检索工具，WebSEEK[五]是一种面向 WWW 的文本或图像搜索引擎。这两个检索系统都是由哥伦比亚大学开发的。它们的主要特点是采用了图像区域之间空间关系和从压缩域中提取的视觉特征。系统所采用的视觉特征是利用颜色集和基于小波变换的纹理特征。VisualSEEK 同时支持基于视觉特征的查询和基于空间关系的查询。WebSEEK 包括三个主要模块：图像/视频采集模块，主题分类和索引模块，查找、浏览和检索模块。相对于其他的多媒体检索系统，VisualSEEK 的优点在于：高效的 Web 图像信息检索，采

[一] http：//wwwqbic. almaden. ibm. com。
[二] http：//www. virage. com/cgi - bin/query - e。
[三] http：//vismod. www. media. mit. edu/vismod/demos/photobook/index. html。
[四] http：//www. ctr. comlumbia. edu/VisualSEEk。
[五] http：//www. ctr. comlumbia. edu/WebSEEk。

用了先进的特征抽取技术，用户界面强大，操作简单，查询途径丰富，输出画面生动且支持用户直接下载信息。而 WebSEEK 本身就是一个独立的万维网可视化编程工具，已经对 650000 幅图像和 10000 个影像片段进行了编目，用户可以使用目录浏览和特征检索方式进行图像检索。

基于内容的图像检索从理论上可以分为三个层次：特征语义，即利用图像的颜色、纹理和形状等低层特征及其组合进行检索；对象语义和空间关系语义，即需要利用导出的特征进行一定的逻辑推理，识别出图像中含有的目标；场景语义，行为语义和情感语义，涉及图像的抽象属性，需要对所描述的目标和场景进行高层语义推理。可以看出，当前大多数成型的图像检索系统都停留在第一个层次，如图 5-9 所示，预先按照某种方法提取出查询图像以及图像库中待检索图像的低层特征（如颜色、纹理、形状），待查询图像的低层特征形成一个特征库，然后把查询图像的特征与特征库中的特征进行匹配，以寻找相似的图像[192]。

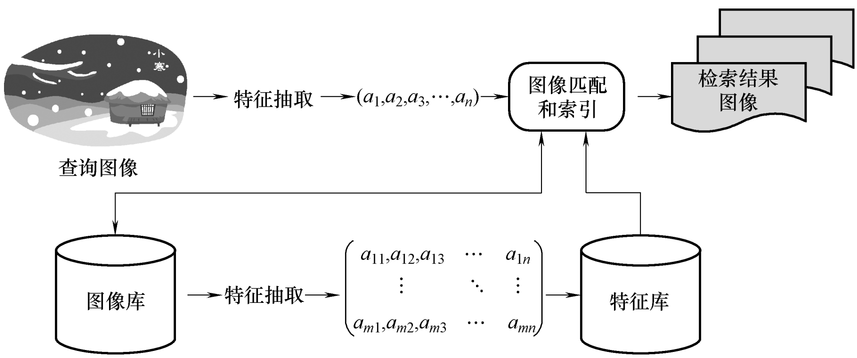


图 5-9 图像检索系统的结构流程（来源：杨红菊，2009 年）

针对以上三个层面，基于内容的图像检索技术研究热点主要可以分为五个方面：

1. 基于整体特征的图像检索

主要研究如何选择合适的图像整体特征描述图像内容和采用什么样的图像度量方法进行图像匹配。由于只是用图像的某些整体特征，不能完整地描述图像的内容，因此图像检索的准确率往往不高。

2. 基于区域的图像检索

主要通过图像分割技术将图像进行划分，然后针对每个区域使用局部特征进行描述，综合区域的局部特征从而得到图像的总体特征，最后使用合适的相

似性度量标准来检索图像。

3. 基于图像语义的研究

相对于图像的颜色、纹理及形状等低层特征而言，语义特征属高层特征，具有主观抽象的特点，是研究的最终目标。目前基于语义特征的图像检索技术的主要研究内容是：如何从多种渠道获取图像语义信息；所获取的语义信息如何与图像低层特征结合；如何通过相关反馈技术在图像之间传递语义信息；以及如何将图像低层特征与图像的关键词结合进行图像的自动标注以提高图像检索的准确率等。

4. 高维索引技术的研究

要想使 CBIR 系统得到实际的应用，那么对于大规模大容量的图像数据库中进行检索要解决的主要问题就是高维特征索引技术。目前提取的特征从几百维到几千维，要在整个数据库中对所有图像进行相似性度量变得不实际。最新的研究模型只能处理几百或几千幅图像，只有这样，在顺序扫描处理这些图像时才不至于严重影响系统的操作性能。目前，在这一研究领域已取得一些进展。例如 K-D 树，R-树、变种 R + 树、R * 树、VA-File 等，但探索更加有效的高维索引技术仍是一个急需解决的问题。

5. 相关反馈技术的研究

该技术基于人机交互的思想，以猜测用户需求为目的，并且根据用户的需求动态调整系统检索时所采用的特征向量或参与检索的不同特征的权重系数，从而尽量缩短减小低层特征和高层语义之间的差距，提高算法的检索结果。相关反馈最先由 Rui Yong 将其由文本检索领域引入到 CBIR 领域，此技术是最近几年 CBIR 研究的热点。为了把用户模型嵌入到图像检索系统，最近几年在 CBIR 领域引入了相关反馈与机器学习机制，将成熟的学习算法与图像检索中的在线学习过程（On-line Learning）结合起来以提高检索准确率。

5.4.2　基于模板匹配的检索方法

作为一种知觉模型，模板匹配（Template Matching）的原理是这样的：我们所遇到并期望从中获得意义的每一个事物、事件或其他刺激，都会与先前已经存储的模式或模板进行比较。因此，知觉的过程包括将输入信息与已经存储的模板进行比较，并从中寻找出一种匹配的模板[13]。如果有一些模型都与之匹配或相近，就需要通过进一步的加工，以区分出哪一个模板是最为合适的。这一模型意味着在我们的知识基础中，已经存储了数以百万计的不同模板——每一

个可以辨识的不同物体或模式，都有一个与之匹配的模板存在。

本书结合局部特征的特点和模板匹配的原理，提出了一种图像检索方法。该方法将从查询图上提取出的每个局部特征都作为单个模板存储起来。对于图像库中的所有图像，都要用前面所述方法判断其每一个局部特征是否和模板之一匹配。如果局部特征和模板匹配的数量越多，则该幅图像和查询图相似的程度就越高。在本书的实验中，使用特征匹配比例（Feature Matching Proportion）来表示相似程度，即

$$Fp = \frac{Mn}{Fn} \tag{5-11}$$

式中，Fp 为特征匹配比例，Mn 为相匹配的局部特征对的数量，Fn 为查询图中局部特征的数量。

由于本书所用的局部特征都能看作高维向量空间中的点，可以通过计算两个点之间的接近程度来衡量图像的局部特征之间的相似度。目前最为常用的相似度度量都具有很强的特征依赖性，不同的特征需要应用不同的度量方法才能获得最佳效果。Mikolajczyk 等人[156]经过大量实验对比，发现对于 SIFT 和 GLOH 等局部特征在图像检索中的应用来说，用欧氏距离作为相似度度量已经可以满足实际应用的要求。

显然，模板匹配并不完全适合知觉原理的实际应用。首先，这一模型要想成立的话，必须存储数量大得令人难以置信的模板；其次，该模型无法解释新的模板是如何创造出来的，又如何保持识别系统与这些数量不断增长的模板的联系；最后，实践中往往会将许多模式或多或少地认为是同样的东西，即使这些模式有比较明显的差别。

5.4.3 基于原型匹配的反馈技术

正如上一小节所述，基于模板匹配的检索方法虽然有很大的潜力和研究空间，但其不足之处也是显而易见的。那就是提取的局部特征描述得过于具体，在检索包含同一个体的图像中效果非常好，却不适于匹配包含某一类物体的图像。比如，从一幅行人图像中得到了一些局部特征，分别描述此人的头、颈、腰、腿，这些特征比较适用于匹配关于该人的其他图像，如果是另外一个人的图像，这些特征就不容易匹配上了，更何况图像库中形体不同、姿态各异的人了。而基于内容的图像检索需要对一类物体进行匹配，比如检索有汽车、飞机、坦克、人群、楼房的图像，这种情况下就需要对具体的局部特征进行组合优化，从而得到对某类物体的理想化表征——原型。

原型匹配理论是这样描述知觉加工的：当一种视觉系统收到一个新刺激，该系统就会将它与原先存储的原型进行比较，但并不要求完全相匹配，事实上大致的匹配就可以了[13]。原型匹配模型允许输入信息与原型之间存在差异，这就赋予了该模型比模板模型更多的灵活性。如图 5-10 所示，我们可以从各式各样的关于人的图像中提取到许多描述人体各个部位的局部特征，然后对这些局部特征集合进行聚类分析，得到这些局部特征的原型特征。

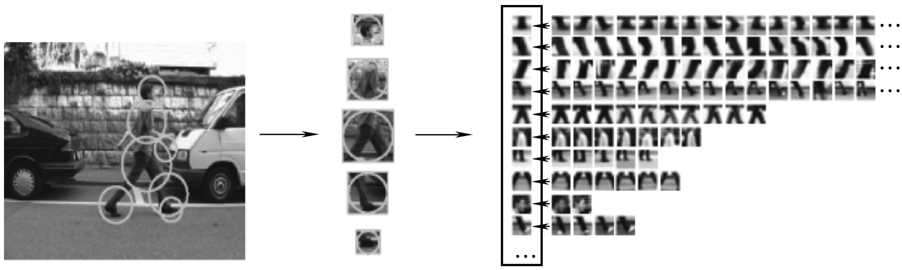


图 5-10　局部特征的原型获取示意图（来源：Leibe，2008 年）

关于聚类分析的算法，本书在第 3 章 3. 5. 1 节进行了介绍，由于图像检索所用到的图像示例比较少，提取的局部特征总数也不是很多。所以划分方法中的 k-平均值（k-means）、k-中心点（k-medoids）和层次方法中的凝聚聚类效果上的差距并不十分明显。

为了实现从模板匹配到原型匹配的转变，我们引入了相关反馈技术。相关反馈技术基于人机交互的思想，以猜测用户需求为目的，并且根据用户的需求动态调整系统检索时所采用的特征向量或参与检索的不同特征的权重系数，从而尽量减小底层特征和高层语义之间的差距，改善算法的检索效果。

在本书中，通过模板匹配的初次检索后，由用户根据自身的信息需求挑选出相关程度较大的检索结果，系统根据用户的反馈进行学习，对这些挑选出的图像以及查询图的局部特征通过上述算法进行组合优化，把得到的“原型”存储起来，此后的处理过程就和基于模板匹配的检索方法类似。

5. 5　实验结果与分析

1. 实验环境

（1）硬件环境

普通 DELL 台式计算机一台，基本配置为 P(R) D/3.4GHz/1.00G/160G/19in（英寸）。

（2）软件环境

WindowsXP 操作系统，Visual Studio C ++ 6.0 开发平台，Matlab2007b，OpenCV 函数库。

2. 实验数据来源

Mikolajczyk 等人构造的图像库㊀可以用于从不同角度对局部特征描述子进行性能测试。该库中含有 8 组（每组 6 幅）PPM 格式的图像，大小为 765 × 512 到 1000 × 700 像素不等，分别代表 5 种不同的图像变换：视点变化（两组图像）、尺度变化（两组图像）、图像噪声（两组图像）、JPEG 压缩、光亮度变化。

本章图像拼接实验所用到的数据是无人机在黄河上空拍摄的凌汛图像序列以及在太原火车站上空拍摄的图像序列，每幅图像皆存储为 JPEG 格式，大小分别为 4727 × 2848 像素和 3888 × 2592 像素。这些可见光图像都来自普通的航拍 CCD 相机或摄像机，传感器设备位于飞机底部的一个近似固定视点，相邻图像间有不小于 16% 的重叠，拍摄所有图像时焦距基本保持不变。

Wan 从 Corel 标准测试图像库中挑选出来的图像被广泛应用于对图像检索的效果验证（见 1.3 节）。如图 5-11 所示，包含非洲原始居民、海滩、建筑物、公交汽车、恐龙、大象、花卉、马、雪山、食品 10 类共计 1000 幅彩色图像，皆存储为 JPEG 格式，大小为 256 × 384 像素或 384 × 256 像素。每一类的 100 幅图像被设定为识别的标准结果。

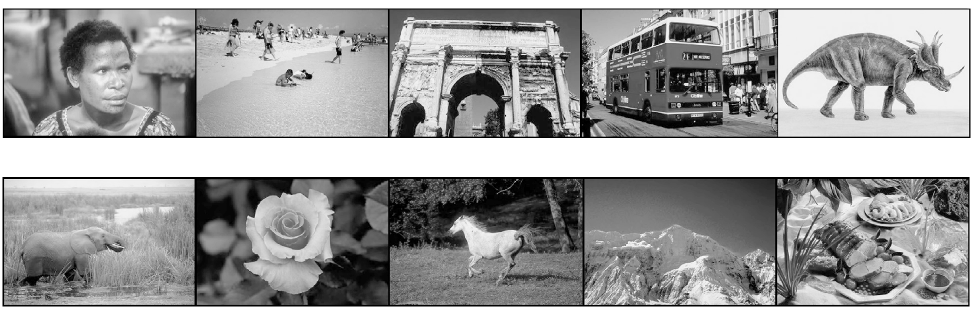


图 5-11　Corel 图像库示例

㊀ http://www.robots.ox.ac.uk/~vgg/research/affine。

实验 1：局部特征在目标匹配中的性能比较

本章实验测试的局部特征有第 2 章介绍的 GLOH、SIFT、PCA-SIFT、SC（Shape Context）、不变矩（Moment Invariants，MI）和导向滤波器（Steerable Filters，SF），所用的衡量性能的标准为 3.4.2 节所提到的查准率（Precision）、查全率（Recall）。图像匹配所用到的图像组在平面内旋转的角度范围是 30°～45°，视点变化的范围是 50°～60°，缩放变化的尺度因子是 2～2.5，最终的实验结果是取各个实验数据的平均值。在匹配策略上，本实验采用 5.2.1 节所提到的最近邻特征和次近邻特征的距离比值，变动该阈值的上限 t，形成了图 5-12 所示的曲线图。

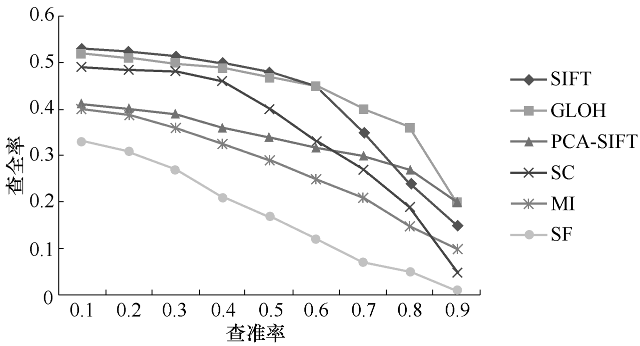


图 5-12　局部特征描述子的性能比较

注意实验中性能评价是建立在对同一物体或场景的匹配识别上的，而且在具体过程中可以发现每种局部特征都有一定的适用范围，例如，SC 描述子在形状特征明显的目标匹配中效果很好，但在纹理图像和非刚性目标的识别中效果不佳。在低维描述子中，不变矩和导向滤波器的性能要略胜一筹。但总体看来，SIFT 和 GLOH 特征的性能最为稳定，应用也比较广泛。

图 5-13 是在不同视点对同一场景进行拍摄的两幅图像，上图是站在地面上的平视拍摄，下图是站在河床底部的仰视拍摄。从匹配效果可以看出，SIFT 特征描述子极大地消除旋转、光照和尺度变化等因素的影响。

实验 2：航拍图像序列拼接

本章实验的目的正是在飞行器和相机具体参数未知的情况下快速拼接航拍图像，不依赖复杂的相机标定设备、旋转台和陀螺仪等；并尽量降低对航拍的限制条件，允许图像之间较大的亮度差异以及相机的轻微晃动等；特殊设备拍摄的照片以及在精确参数下的图像拼接不在本书的研究之列。

图 5-14a 是无人机航拍的黄河凌汛的一组照片，图像上主要是自然景物地

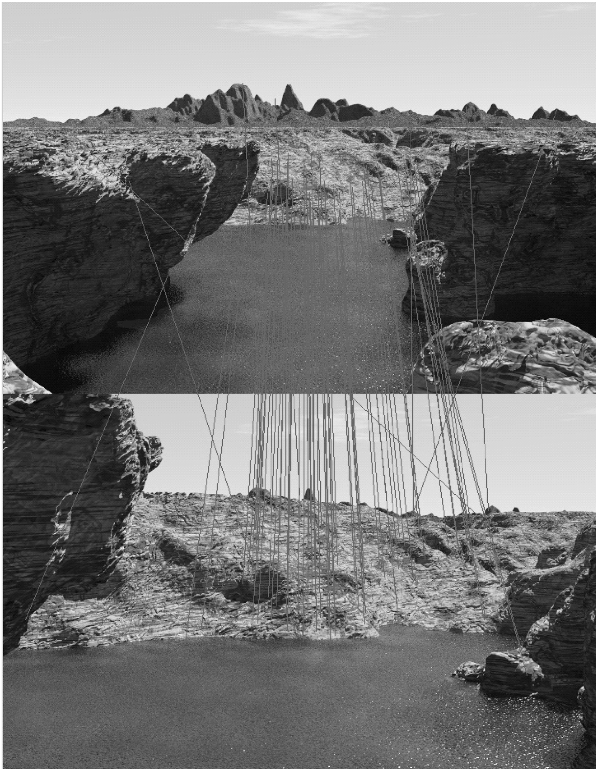
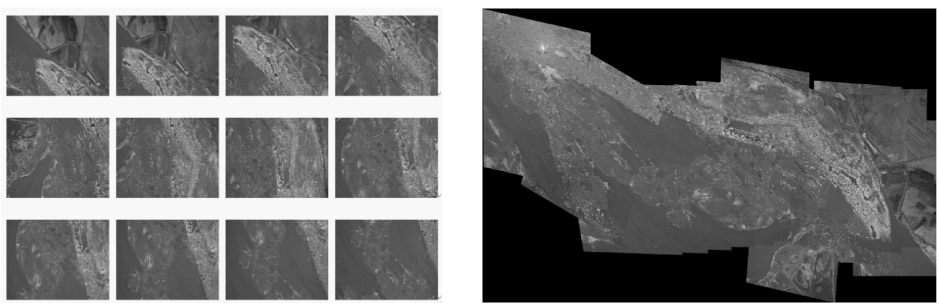


图 5-13 利用 SIFT 特征进行目标匹配

貌，人造目标比较少，这对于计算机自动拼接是一个挑战。但如图 5-14b 所示，本书利用 SIFT 特征进行拼接方法十分稳健，局部特征提取技术减少了噪声干扰和光照变化的影响；多分辨率技术的应用也有效地降低了图像配准的计算开销；

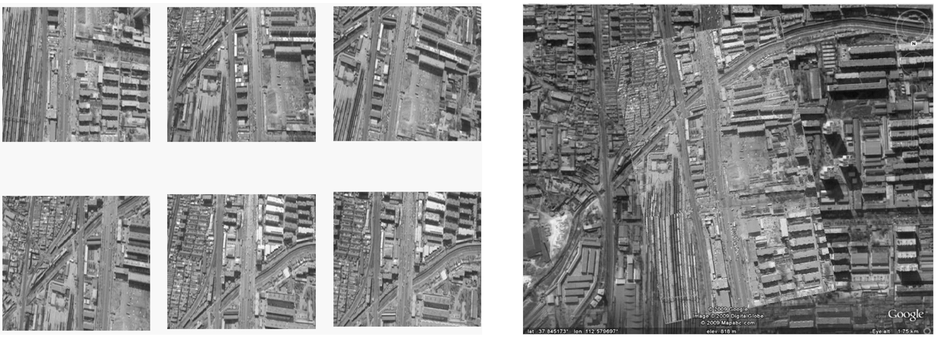


a)　　b)

图 5-14 黄河凌汛的航拍图像拼接结果

a）关于黄河凌汛的航拍图像序列　b）拼接后的效果

通过比较最近邻点和次近邻点的距离的方法也可以有效地剔除“外点”。



a)　　b)

图 5-15　太原火车站的航拍图像拼接结果

a）太原火车站的航拍图像序列　b）拼接后嵌入地图的效果

图 5-15a 所示为无人机在太原火车站上空拍摄的一组照片，图 5-15b 所示为其拼接后嵌入地图的效果。可以看出，计算机自动拼接后的航拍图像在严格对应于地理信息系统时还存在一些问题。但从一般意义上讲，这部分内容并不属于图像自动拼接技术的研究范畴，可以在下一步工作中引入人机交互的方法，根据相关参数对拼接图像进行几何校正。

本书提出的方法也适用于航拍视频图像拼接，图 5-16 所示为从一段空中鸟瞰城市的视频里抽取图像进行准实时拼接的效果。这也表明本书介绍的方法稳定、可靠，在保证运算速度的同时依然能够取得很好的视觉效果。



图 5-16　航拍视频图像拼接结果

实验 3：基于局部特征的图像检索

近年来，可伸缩颜色描述符（Scalable Color Descriptor，SCD）、基于曲率尺度空间（Curvature Scale Space，CSS）的形状描述符、欧拉向量（EulerXor）广泛应用于基于内容的图像检索领域。SCD 是 MPEG-7 推荐的四个可以独立运用的颜色描述符之一，CSS 也是 MPEG-7 指定的形状描述方法，EulerXor 是灰度图像的组合特征，它们都具有维数小、计算简便、对平移和旋转不敏感的特点。本章将本书提出的图像检索方法与利用以上四种特征的检索方法进行对比，实验结果如图 5-17 所示。

由于 Corel 图像库中每类图像的颜色特征比较明显，对类别的区分度较高，SCD 的效果非常好；相对而言，利用 GLOH 特征的模板匹配方法（GLOH-TM）效果最差，这是因为匹配方法过于简单，没有对特征进行相应的处理；而通过对局部特征进行组合优化，利用 GLOH 特征的原型匹配方法（GLOH-PM）的检索效果和 SCD 差距很小，整体表现比较稳定。

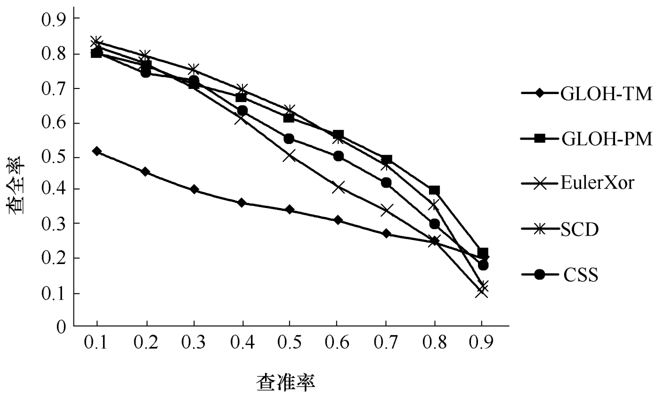


图 5-17 五种图像检索方法的性能比较

5.6 本章小结

局部特征的提出使得目标匹配可以从整体匹配的形式转变为局部匹配的形式，这就为复杂背景下的图像目标识别提供了一条有效的途径。本章在对国内外相关领域的众多研究成果进行深入探讨之后提出了一种局部特征的匹配算法，该算法使用 BBF 算法进行邻近点搜索，通过最邻近距离比值甄别错误匹配，并结合霍夫变换的思想进行遮挡目标的匹配识别。

针对目前图像拼接和图像检索方法的不足之处，本章结合上述匹配算法提出了基于多分辨率技术的航拍图像拼接方法，以及基于原型匹配的图像检索方法。与当前的主流方法相比，进一步验证了利用局部特征进行图像拼接和检索的可行性。在下一步工作中，如果能将局部特征与颜色、纹理、空间关系等整体特征结合起来，会有更好的实用价值和发展前景。

第 6 章　基于局部特征的目标分类

概念和分类是人类思考和行为的建筑基石。

6.1　引言

近些年来，目标分类识别的应用范围越加广泛，成为图像信息处理领域的一个研究热点。其理论方法主要采用无结构的特征组织方式，目的旨在通过训练分类器或特定的网络结构，完成对特定的特征空间中点的划分，形成某些具有相似特性的点的集合。分类训练方法主要是自底向上由数据驱动，通过对训练样本的监督学习，在样本空间产生合适的区分函数，采用形成的分类器或结构参数对待识别目标进行分类决策得到最终的目标识别结果。

判别分类方法可以从图像数据中获取简单的结构化语义，进一步体现目标之间的关系，但是如 4.1 节所述，整体特征自身的局限性极大影响了图像目标分类的实际效果。而由于局部特征性能相对优越，其含有的局部信息可以对图像的内容进行多语义层次的描述，不少研究人员也在尝试将其应用于目标分类[78,126,127,193]。

本书在 1.2.3 节中曾经论述过，人类进行认知过程时，大脑皮层间不仅存在着上行的前馈投射（提示层次性整合），还存在着大量从高级皮层向初级皮层

的反馈投射（提示整体性调节）。层次性与整体性两种机制在大脑皮层的认知过程中是密不可分、难以割裂的。一个值得注意的现象就是在认知过程中存在“局部－整体效应”和“组合效应”。例如，在人脸识别过程中，目标往往会被表达成一个不可分割的整体，而不是仅仅简单的局部组合。

局部特征的单纯组合或者全局尺度的特征（整体特征）虽然能够在一定程度上对图像目标进行描述和区分，但是对目标发生的某些局部或特定场景的变化（比如目标遮挡、背景干扰以及光照、角度、仿射变换等）敏感，其稳定性和可区分性不高。人类思维中的概念既有较简单的基本概念（低层语义，语义粒度最细），也有抽象程度较高的概念（较高层语义，语义粒度较粗），各层次语义使得人的思维中语义概念具有丰富语义粒度。从人类认知角度看，分析理解的过程是不同层次、不同粒度语义信息的交互过程。

在这里，对第一章中图 1-1 所示的图像目标识别系统的基本框架进行相应的改进，可以提出一种结合两种思路（见本书 1. 2. 3 节）的目标识别模型，该模型的建立过程如图 6-1 所示。其核心思想就是从训练样本中提取出多层次的目标特征，然后利用机器学习的方法学习获取的语义概念并建立起有效的知识模型，最终就可以在知识模型的指导下进行测试图像的信息处理和分类识别。

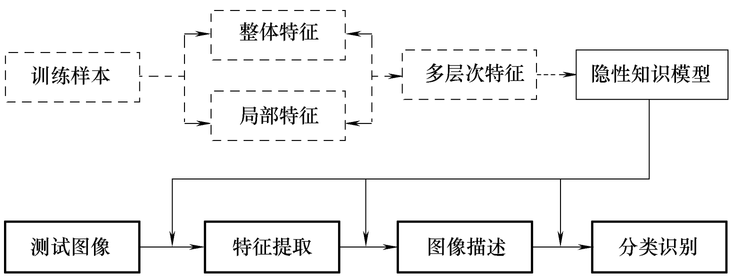


图 6-1　知识驱动的图像识别模型

通过上述分析可以看出，要使计算机能够高效地处理真实图像并对图像中的目标进行识别，建立图像认知的层次性与整体机制是十分必要的。这也就意味着我们必须找到一种理想的形式化表示方法，这种表示一方面要能够真实地简约地反映图像目标的内容，另一方面，要有对不同图像目标的区分能力。目前基于局部特征的图像目标表示通常分为三类，分别是向量空间模型（Vector Space Model，VSM）、滑动窗口模型和结构关系模型。其中，向量空间模型表达简洁、应用方便，不用考虑特征项之间的空间关系，是当前图像目标分类的主流方法。

6.2　目标的向量空间模型表示

向量空间模型，又称特征包模型或词袋模型，是 Salton 等人[194]在 20 世纪 70 年代初提出的，最早用在 SMART 信息检索系统中，此后逐渐发展成为自然语言处理中常用的模型，近几年也被广泛应用在图像目标识别中。

下面给出 VSM 应用在图像识别领域的一些概念。

- 目标（Target）：也称对象或物体，通常是图像中具有某种相似属性的同质区域，如图像分割产生的子区域、客观存在的具有某种物理或语义意义的实体直至整幅图像，参见图 2-3，在本章节的论述中，对目标和图像的概念不加区分。

- 项/特征项（Term/Feature Term）：特征项是 VSM 中最小的不可分的语义单元，可以是任意分割程度上的子区域。一个目标的内容被看成它含有的特征项所组成的集合，表示为 Target = $\boldsymbol{T}(t_1, t_2, \cdots, t_n)$，其中 t_k是特征项，$1 \leqslant k \leqslant n$。

- 项的权重（Term Weight）：对于含有 n 个特征项的目标 $T(t_1, t_2, \cdots, t_n)$，每一特征项 t_k 都依据一定的原则被赋予一个权重 w_k，表示它们在目标描述中的重要程度。这样一个目标 T 可用它含有的特征项及其特征项所对应的权重所表示：$\boldsymbol{T} = \boldsymbol{T}(t_1, w_1; t_2, w_2; \cdots; t_n, w_n)$，简记为 $\boldsymbol{T} = \boldsymbol{T}(w_1, w_2, \cdots, w_n)$，其中 w_k就是特征项 t_k 的权重，$1 \leqslant k \leqslant n$。

一个目标在上述约定下可以看成是 n 维空间中的一个向量，这就是向量空间模型的由来。下面结合目标的表示，给出其定义。

定义 6-1（向量空间模型）　给定一个目标 $\boldsymbol{T}(t_1, w_1; t_2, w_2; \cdots; t_n, w_n)$，$\boldsymbol{T}$ 符合以下两条约定：

1）各个特征项 $t_k(1 \leqslant k \leqslant n)$互异（即没有重复）；

2）各个特征项 t_k无先后顺序关系（即不考虑目标的内部结构）。

在以上两个约定下，可以把特征项 $t_1, t_2, \cdots, t_n$看成一个 n 维坐标系，而权重 $w_1, w_2, \cdots, w_n$为相应的坐标值，因此，一个目标就表示为 n 维空间中的一个向量。我们称 $\boldsymbol{T} = \boldsymbol{T}(w_1, w_2, \cdots, w_n)$ 为目标 T 的向量表示或向量空间模型，如图 6-2 所示。

定义 6-2（向量的相似度度量）　任意两个目标 $\boldsymbol{T}_1$ 和 $\boldsymbol{T}_2$之间的相似系数 $Sim(\boldsymbol{T}_1, \boldsymbol{T}_2)$指两个目标内容的相关程度（Degree of Relevance）。设目标 $\boldsymbol{T}_1$ 和 $\boldsymbol{T}_2$ 表示 VSM 中的两个向量：

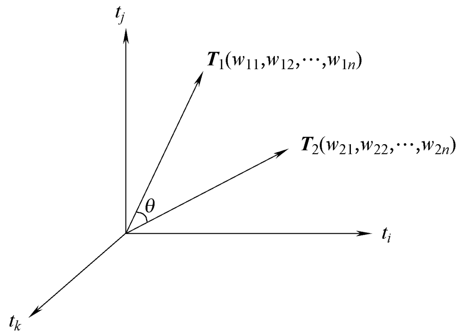


图 6-2　目标的向量空间模型示意图

$$\begin{aligned}\boldsymbol{T}_1 &= \boldsymbol{T}_1(w_{11}, w_{12}, \cdots, w_{1n})\\ \boldsymbol{T}_2 &= \boldsymbol{T}_2(w_{21}, w_{22}, \cdots, w_{2n})\end{aligned} \tag{6-1}$$

那么，可以借助 n 维空间中两个向量之间的某种距离来表示目标间的相似系数，常用的方法是使用向量之间的内积[129]来计算：

$$Sim(\boldsymbol{T}_1, \boldsymbol{T}_2) = \sum_{k=1}^{n} w_{1k} \times w_{2k} \tag{6-2}$$

如果考虑向量的归一化，则可使用两个向量夹角的余弦值来表示相似系数：

$$Sim(\boldsymbol{T}_1, \boldsymbol{T}_2) = \cos\theta = \frac{\sum_{k=1}^{n} w_{1k} \times w_{2k}}{\sqrt{\left(\sum_{k=1}^{n} w_{1k}^2\right)\left(\sum_{k=1}^{n} w_{2k}^2\right)}} \tag{6-3}$$

采用向量空间模型进行目标表示时，需要经过以下两个主要步骤：

1）根据训练样本生成目标表示所需要的特征项序列 $T=\{t_1, t_2, \cdots, t_d\}$；

2）依据目标特征项序列，对训练集和测试集中的各个目标样本进行权重赋值、规范化等处理，将其转化为机器学习算法所需的模式向量。

6.3　构造视觉单词库

在目标识别领域，向量空间模型之所以称为特征包模型或词袋模型，是因为它将目标图像看成由大量的视觉单词（Visual Word）构成。如图 6-3 所示，在目标分类识别中，目标的类别相当于文档的主题，而文档的主题通过其词句来判断，同样，某个目标的类别可以通过构成它的视觉单词进行决策。

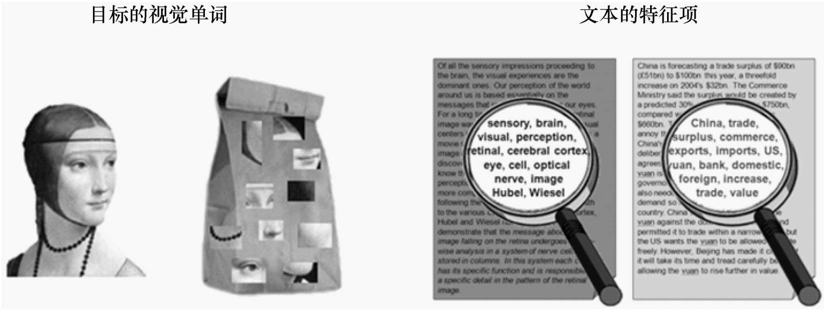


图 6-3　特征包模型示意图

6.3.1　视觉单词的生成方法

从大量样本中提取的局部特征千差万别且数量巨大，如图 6-4 所示，哪怕是从同类目标上提取的描述同一部件的局部特征，也往往是有所不同。这些“模板”描述得过于具体，虽然可以对某一个体进行精确匹配，但不适于对一类目标的识别。这就需要像文本中的词句一样，从众多具体事物的描述中抽象出“概念”，从而抓住一类事物的共性。比如，我们生活中见过许多狗，当提及“狗”的时候，我们想到的应该不是某个特定的狗，而是狗的理想化模型——是对一个非常典型的狗的描述，与其完全相像的狗在现实生活中也许存在，也许根本不存在。

正如本书 5.4 节所述，相近的局部特征经过优化组合之后可以形成“原型”特征，也就是视觉单词。大量的视觉单词就组成了视觉单词库，在一些文献中也称之为码书（Codebook）。用视觉单词作为向量空间模型中的特征项，就可以解决目标图像的表示问题，从而实现基于向量空间模型的目标分类了。

对局部特征进行聚类是构造视觉单词的一种有效途径，因为聚类分析的目的就是将物理或抽象对象的集合分组成由类似的对象组成的多个类[128]。5.4 节简单介绍了聚类算法的几种类型，其中最为常用的是划分方法中的 k-平均值（k-means）和层次方法中的凝聚（Agglomerative）聚类。

k-means 算法是根据预定的类别数目 k 随机地选取 k 个对象作为初始的簇中心。对剩余的每个对象，根据其与各个簇中心的距离，将它赋给最近的簇。然后重新计算每个簇的平均值。这个过程不断重复，直到准则函数收敛。对处理大数据集，该算法是相对可伸缩和高效率的，它的复杂度是 $O(nkt)$，其中，n 是所有对象的数目，k 是簇的数目，t 是迭代的次数，通常 $k \ll n$，且 $t \ll n$。但是

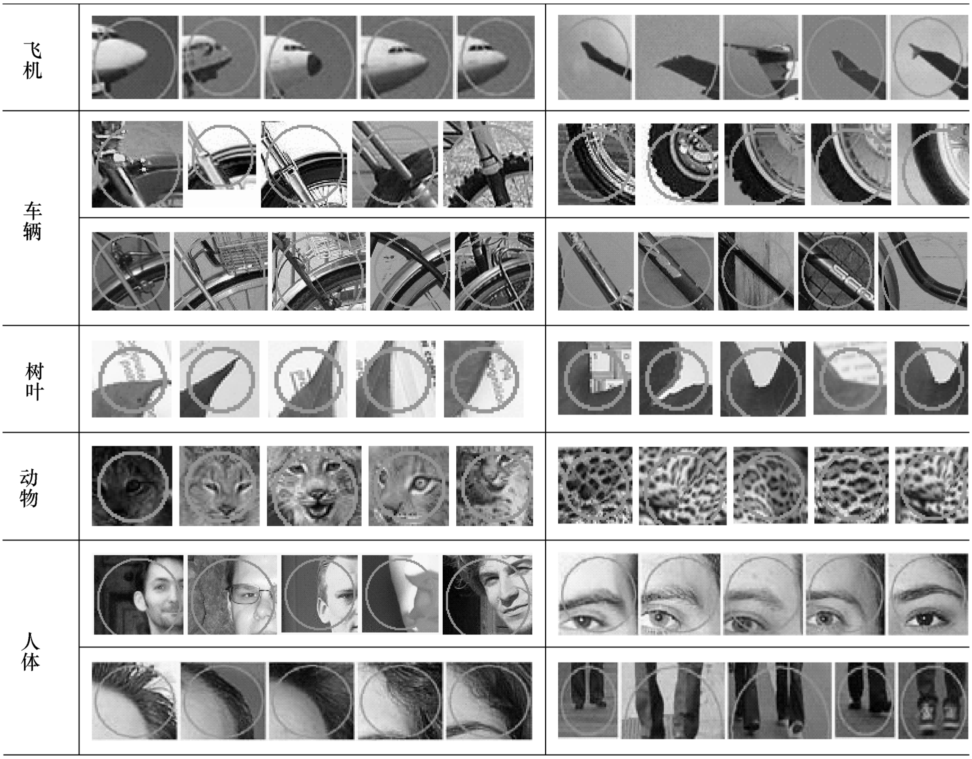


图 6-4　各种局部特征示例

这个算法经常以局部最优结束，必须实现给出簇的数目，对“噪声”和孤立点数据非常敏感，而且不适合与发现非凸面形状的簇。

凝聚聚类是将每个对象作为一个簇，然后合并这些原子簇为越来越大的簇，直到所有对象都在一个簇中，或者某个终结条件被满足。然而，凝聚聚类尽管简单，但经常会遇到合并点的选择困难。这样的决定非常关键，因为一旦一组对象被合并，下一步的处理将在新生成的簇上进行，这一步骤无法撤销，聚类之间也不能交换对象。所以，每次合并之前需要检查和估算大量的对象或簇，其过高的时间复杂度和空间复杂度严重制约了该算法的应用。

6.3.2　基于 RNN 的层次聚类算法

RNN 算法对标准层次聚类的合并准则和相似度度量做了相应的改进，从而降低了其复杂度，使其更适用于大规模的数据集。该算法的基本原理虽然在 20 多年前已经提出，但是直到最近几年才被应用在目标识别领域[127]。RNN 算法的核心思想是构造相互最近邻对（Reciprocal Nearest Neighbor Pairs，RNNP），也就

是一对互为最近邻的数据点，这就满足了聚类的可还原性——当两个簇 C_i 和 C_j 进行合并之后，相对其他任意簇 C_k 的相似度要减小，其表达式如下：

$$sim(C_i, C_j) \geqslant \sup(sim(C_i, C_k), sim(C_j, C_k)) \Rightarrow$$
$$\sup(sim(C_i, C_k), sim(C_j, C_k)) \geqslant sim(C_i \cup C_j, C_k) \tag{6-4}$$

这就保证了合并最近邻对时不改变与任何其他簇的最邻近关系，而且该性质对于平均距离和平均值的距离两种簇间距离度量方法来说，都是完备的。令 $X=\{x^{(1)}, \cdots, x^{(N)}\}$ 和 $Y=\{y^{(1)}, \cdots, y^{(M)}\}$ 为两个簇，则这两种簇间距离度量的定义为

平均距离：$$sim(X, Y)=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} sim(x^{(i)}, y^{(j)}) \tag{6-5}$$

平均值的距离：$$sim(X, Y) = sim\left(\frac{1}{N}\sum_{i=1}^{N}x^{(i)}, \frac{1}{M}\sum_{j=1}^{M}y^{(j)}\right) \tag{6-6}$$

本书采用平均距离作为两个簇的相似度度量，在聚类过程中为任意一个数据点建立一条最近邻链（Nearest Neighbor Chain），通过最近邻链来简单有效地寻找到最近邻对。具体步骤如下：

算法：基于最近邻链的 RNN 凝聚聚类

```
//随机选定一个数据点 v∈V 初始化链表 L
//剩余的数据点都包含在集合 R 中
last←0; lastsim [0] ←0
L [last] ←v∈V; R←V \ v
While R≠∅ do
    //在集合 R 中搜索下一个最近邻点并计算相似度
    (s, sim) ←getNearestNeighbor (L [last], R)
  if sim > lastsim [last] then
    //没有找到最近邻对，把 s 添加到最近邻链中
    last←last + 1
    L [last] ←s; R←R \ {s}
    lastsim [last] ←sim
  else
    //找到最近邻对，合并链表最后两个节点
    if lastsim [last] > t then
      s←agglomerate (L [last], L [last - 1])
      R←R∪ {s}
      last←last - 2
  else
```

（续）

```
      //丢弃当前链表
      last← -1
    end if
  end if
  if last <0 then
      //重新随机选择一个数据点 v∈R 建立一个新链表
      last←last +1
      L [last] ←v∈R; R←R\ {v}
    end if
  end while
```

整个聚类过程需要 $3(N-1)$次迭代，其搜索最近邻点的时间代价最低可以降到 $O(n)$。当合并最近邻对得到一个新的簇时，需要重新计算该簇与其他各个簇的相似度，如果通过平均值的距离来度量两个簇的距离，其计算复杂度仅为 $O(n)$,但是由于本书采用的是平均距离，则需要通过更为有效的方法进一步降低复杂度。

设 μ_x，μ_y 和 σ_x^2，σ_y^2 分别为簇 X，Y 的平均值和方差，两个簇的平均距离（在欧氏空间中）可以用下面的公式表示：

$$sim(X, Y) = -\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(x^{(i)} - y^{(j)})^2 = -(\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2) \quad (6\text{-}7)$$

采用这种形式来计算簇间距离，只需要储存每个簇的平均值和方差。当合并最近邻对产生新簇的时候，新簇的平均值和方差计算公式如下：

$$\mu_{new} = \frac{N\mu_x + M\mu_y}{N + M} \quad (6\text{-}8)$$

$$\sigma_{new}^2 = \frac{1}{N + M}\left(N\sigma_x^2 + M\sigma_y^2 + \frac{NM}{N + M}(\mu_x, \mu_y)^2\right) \quad (6\text{-}9)$$

如此一来，算法的时间复杂度为 $O(n^2)$，空间复杂度为 $O(n)$。对于低维数据，还可以通过更为有效的最近邻搜索技术进一步降低复杂度。

6.4 基于信息论的特征选择方法

解决“维数灾难”现象是模式识别领域的一个非常重要的任务，因为提取出的原始特征往往数量庞大，不仅增加了计算复杂度，而且很大程度上影响了

分类器的设计及其性能。这就需要从一组特征中挑选出一些最有效的特征以达到降低特征空间维数的目的，这个过程叫做特征选择或特征压缩。

最简单的特征选择方法是根据专家（相关领域的科研人员）的知识挑选出那些对分类识别最有影响的特征；另一个可能则是用统计学和信息论的方法进行筛选比较，来找出最有分类信息的特征。显然，前者受到太多的条件限制不是很实用，而后者则是当前模式识别领域的研究热点。

目前已有的特征选择方法比较多，其中基于图像频率的特征选择方法简单易行，可以在降低特征空间复杂度的同时去掉一部分噪声特征，但低频特征也可能带有很大的信息量，该方法直接去除低频特征会影响识别效果；χ^2 统计量度量特征和类别独立性的缺乏程度，优点是降维效果比较好，缺点则是统计花费大；术语强度的特点是基于目标聚类的方法，认为在相关目标中出现次数越多的特征具有信息量，这样可以去掉大部分无信息量或带有很少信息量的特征，但在图像目标分类的实验中效果不是很好。本书的 2.4.1 节对这些方法的特点也做了相应评述，基于这些分析以及目标分类的具体应用特点，本章分别采用了信息论中的信息增益（IG）法和互信息（MI）法对图像特征进行筛选。

6.4.1　信息论的相关概念

信息是个相当宽泛的概念，很难用一个简单的定义将其完全准确地把握。然而对于任何一个概率分布，可以定义一个称为熵（Entropy）的量，它具有许多特性符合度量信息的直观要求，是信息论的基本概念。

如果 X 是一个离散型随机变量，取值空间为 R，其概率分布为 $p(x)=P(X=x)$，$x\in R$。那么 X 的熵 $H(X)$定义为

$$H(X) = -\sum_{x\in R} p(x)\log_2 p(x) \tag{6-10}$$

有时也将 $H(X)$记为 $H(p)$。其中对数以 2 为底，熵的单位用比特（二进制位）表示。所以通常将 $\log_2 p(x)$简写成 $\log p(x)$，并约定 $0\log 0=0$。

熵又称为自信息（Self-information），可以视为描述一个随机变量的不确定性的数量。它表示信源 X 每发出一个符号（不论发出什么符号）所提供的平均信息量[81]。一个随机变量的熵越大，它的不确定性越大，那么，正确估计其值的可能性就越小。越不确定的随机变量越需要大的信息量用以确定其值。

如果 X，Y 是一对离散型随机变量，X，$Y\sim p(x, y)$，X，Y 的联合熵（Joint Entropy）$H(X, Y)$定义为

$$H(X, Y) = -\sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(x, y) \tag{6-11}$$

联合熵实际上就是描述一对随机变量平均所需要的信息量。

给定随机变量 X 的情况下，随机变量 Y 的条件熵（Conditional Entropy）定义如下：

$$\begin{aligned} H(Y \mid X) &= \sum_{x\in X} p(x) H(Y \mid X = x) \\ &= \sum_{x\in X} p(x)\Big[-\sum_{y\in Y} p(y \mid x)\log p(y \mid x)\Big] \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(y \mid x) \end{aligned} \tag{6-12}$$

将式（6-11）中的联合概率 $p(x, y)$展开，可得

$$\begin{aligned} H(X, Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x, y)\log[p(x)p(y \mid x)] \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x, y)[\log p(x) + \log p(y \mid x)] \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(x) - \sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(y \mid x) \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(x) - \sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(y \mid x) \\ &= H(X) + H(Y \mid X) \end{aligned} \tag{6-13}$$

该式称为熵的连锁规则（Chain Rule of Entropy）。推广到一般情况，有

$$H(X_1, X_2, \cdots, X_n) = H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_n \mid X_1, \cdots, X_{n-1}) \tag{6-14}$$

6.4.2　基于信息增益法的特征选择

利用信息增益法选择特征，是依据某个特征项 t_i为整个分类所能提供的信息量多少来衡量该特征项的重要程度，从而决定对该特征项的取舍。某个特征项 t_i 的信息增益是指有该特征或没有该特征时，为整个分类所能提供的信息量的差别，其中，信息量的多少就用熵来衡量。可以计算出不考虑任何特征时目标的熵以及考虑该特征后目标的熵，并将两者之间的差值定义为信息增益：

$$\begin{aligned} IG(t_i) &= H(T) - H(T \mid t_i) \\ &= \Big\{-\sum_{j=1}^{M} P(C_j)\times\log P(C_j)\Big\} - \Big\{P(t_i)\times\Big[-\sum_{j=1}^{M} P(C_j \mid t_i)\times\log P(C_j \mid t_i)\Big] \\ &\quad + P(\overline{t_i})\times\Big[-\sum_{j=1}^{M} P(C_j \mid \overline{t_i})\times\log P(C_j \mid \overline{t_i})\Big]\Big\} \end{aligned} \tag{6-15}$$

其中，$P(C_j)$ 表示 C_j类目标在样本集中出现的概率，$P(t_i)$ 表示样本集中包含

特征项 t_i 的目标的概率，$P(C_j \mid t_i)$ 表示目标包含特征项 t_i 时属于 C_j 类的条件概率，$P(\overline{t_i})$ 表示样本集中不包含特征项 t_i 的目标的概率，$P(C_j \mid \overline{t_i})$ 表示目标不包含特征项 t_i 时属于 C_j 类的条件概率，M 表示类别数。

从信息增益的定义可知，一个特征的信息增益实际上描述的是它包含的能够帮助预测类别属性的信息量。从理论上讲，信息增益应该是最好的特征选择方法，但实际上由于许多信息增益比较高的特征出现频率往往较低，所以当使用信息增益选择的特征数目比较少时，往往会存在数据稀疏问题，此时识别效果也比较差。对此的改进方法是，首先对训练集中出现的每个特征项计算其信息增益，然后指定一个阈值，从特征空间中移除那些信息增益低于此阈值的特征项；或者指定保留的特征项个数，按照增益值从高到低的顺序选择特征项组成特征向量。

6.4.3　基于 CHI 统计量的特征选择

χ^2 统计量（CHI）衡量的是特征项 t_i 和类别 C_j 之间的相关联程度，并假设 t_i 和 C_j 之间符合具有一阶自由度的 χ^2 分布。特征对于某类的 χ^2 统计值越高，它与该类之间的相关性越大，携带的类别信息也越多，反之则越少。

如果令 N 表示训练集中样本的总数，A 表示属于 C_j 类且包含 t_i 的目标频数，B 表示不属于 C_j 类但包含 t_i 的目标频数，C 表示属于 C_j 类但不包含 t_i 的目标频数，D 是既不属于 C_j 类也不包含 t_i 的目标频数。上述四种情况可以用表 6-1 表示。

表 6-1　特征项与类别关系的表示

类别 / 特征项	C_j	$\sim C_j$
t_i	A	B
$\sim t_i$	C	D

特征项 t_i 对 C_j 的 CHI 值为

$$\chi^2(t_i, C_j) = \frac{N \times (A \times D - C \times B)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)} \tag{6-16}$$

对于多类问题，基于 CHI 统计量的特征选择方法可以采用两种实现方法：一种是分别计算 t_i 对于每个类别的 CHI 值，然后在整个训练集上计算：

$$\chi^2_{\mathrm{MAX}}(t_i) = \max_{j=1}^{M}\{\chi^2(t_i, C_j)\} \tag{6-17}$$

式中，M 为类别数。从原始特征空间中去除统计量低于给定阈值的特征，保留统计量高于给定阈值的特征作为目标特征。另一种方法是，计算各特征对于各

类别的平均值：

$$\chi^2_{\mathrm{AVG}}(t_i) = \sum_{j=1}^{M} P(C_j)\chi^2(t_i, C_j) \tag{6-18}$$

以这个平均值作为各类别的 CHI 值。但有研究表明，后一种方法的表现不如前一种方法。

6.4.4　基于互信息法的特征选择

根据熵的连锁规则，有

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{6-19}$$

因此，

$$H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{6-20}$$

这个差叫做 X 和 Y 的互信息，记作 $I(X;Y)$。或者定义为：如果 $(X, Y) \sim p(x, y)$，则 X，Y 之间的互信息 $I(X;Y) = H(X) - H(X|Y)$。

互信息是一个均衡非负的信息测度，$I(X;Y)$反映的是在知道了 Y 的值以后 X 的不确定性的减少量。可以理解为 Y 的值透漏了多少关于 X 的信息量。互信息和熵之间的关系可以用图 6-5 表示。

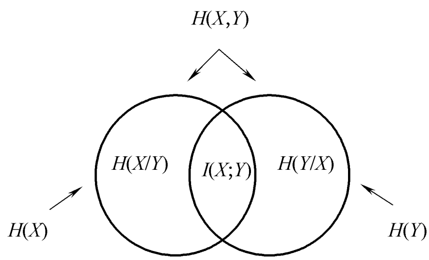


图 6-5　互信息和熵之间的关系示意图

如果将定义中的 $H(X)$和 $H(X|Y)$展开，可得

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X, Y) \\ &= \sum_x p(x)\log\frac{1}{p(x)} + \sum_y p(y)\log\frac{1}{p(y)} + \sum_{x,y} p(x, y)\log p(x, y) \\ &= \sum_{x,y} p(x, y)\log\frac{p(x, y)}{p(x)p(y)} \end{aligned} \tag{6-21}$$

由于 $H(X|X) = 0$，因此

$$H(X) = H(X) - H(X|X) = I(X;X) \tag{6-22}$$

这一方面说明了熵之所以称为自信息的原因，另一方面说明了两个完全相互依赖的变量之间的互信息并不是一个常量，而是取决于它们的熵。实际上，互信息体现了两变量之间的依赖程度：如果 $I(X;Y)>>0$，表明 X 和 Y 是高度相关的；如果 $I(X;Y)=0$，表明 X 和 Y 是相互独立的；如果 $I(X;Y)<<0$，表明 X 和 Y 是互不相关的分布。

利用互信息法选择特征的基本原则是选择类别相关的特征，同时排除冗余的特征。特征与类别之间的互信息很好地度量了特征的相关性，特征与特征之间的互信息则度量它们之间的相似性（冗余性）。因此，基于互信息的特征选择一般遵循这样一种模式：在顺序前向搜索中寻找与类别互信息最大而与前面已选特征互信息最小的特征项。

在目标分类中可以简单认为：互信息越大，特征 t_i 和类别 C_j 共现的程度越大。那么，t_i 和 C_j 的互信息可以由下式计算：

$$I(t_i, C_j)=\log\frac{P(t_i, C_j)}{P(t_i)P(C_j)}=\log\frac{P(t_i|C_j)}{P(t_i)}\approx\log\frac{A\times N}{(A+C)\times(A+B)} \tag{6-23}$$

式中，A、B、C、D 的含义和 6.4.3 节中约定的完全相同。如果特征 t_i 和类别 C_j 无关，则 $P(t_i, C_j)=P(t_i)\times P(C_j)$，那么，$I(t_i, C_j)=0$。

为了选出对多类图像目标识别有用的特征，与上面的基于 CHI 统计量的处理方法类似，基于互信息法的特征选择也可以采用最大值和平均值两种实现方法：

$$I_{\text{MAX}}(t_i)=\max_{j=1}^{M}[P(C_j)\times I(t_i, C_j)] \tag{6-24}$$

$$I_{\text{AVG}}(t_i)=\sum_{j=1}^{M}P(C_j)I(t_i, C_j) \tag{6-25}$$

6.5　视觉单词的权重计算

视觉单词权重用于衡量某个视觉单词（特征项）在目标表示中的重要程度或者区分能力的强弱。权重计算的一般方法是利用训练集样本的统计信息，主要是词频，给视觉单词赋予一定的权重。注意，“词频”以及后面提到的“文档频度”，都是在文本分类中产生的，在本章节中用图像目标相关的概念进行理解即可，不再特意进行替换。

本书参阅相关文献，将一些常用的权重计算方法归纳为表 6-2 所示的形式。表中各变量的说明如下：w_{ij} 表示特征项 t_i 在目标 T_j 中的权重，tf_{ij} 表示特征项 t_i 在

训练样本 T_j 中出现的频度；n_i 是训练集中出现特征项 t_i 的样本数，N 是训练集中总共的样本数；M 为特征项的个数，nt_i 为特征项 t_i 在训练样本中出现的次数。

表 6-2 特征权重的计算方法

名称	权重函数	说明
布尔权重	$w_{ij}=\begin{cases}1，如果\ tf_{ij}>0\\0，否则\end{cases}$	如果目标中出现该特征项，那么模式向量的该分量为 1，否则为 0
绝对词频（TF）	tf_{ij}	使用特征项在目标中出现的频度表示目标
倒排文档频度（IDF）	$w_{ij}=\log\frac{N}{n_i}$	稀有特征比常用特征含有更新的信息
TF-IDF	$w_{ij}=tf_{ij}\times\log\frac{N}{n_i}$	权重与特征项在目标中出现的频率成正比，与在整个训练集中出现该特征项的样本数成反比
TFC	$w_{ij}=\frac{tf_{ij}\times\log(N/n_i)}{\sqrt{\sum_{t_i\in T_j}[tf_{ij}\times\log(N/n_i)]^2}}$	对目标长度进行归一化处理后的 TF-IDF
ITC	$w_{ij}=\frac{\log(tf_{ij}+1.0)\times\log(N/n_i)}{\sqrt{\sum_{t_i\in T_j}[\log(tf_{ij}+1.0)\times\log(N/n_i)]^2}}$	在 TFC 基础上，用 tf_{ij} 的对数值代替 tf_{ij} 值
熵权重	$w_{ij}=\log(tf_{ij}+1.0)\times\left(1+\frac{1}{\log N}\sum_{j=1}^{N}\left[\frac{tf_{ij}}{n_i}\log\left(\frac{tf_{ij}}{n_i}\right)\right]\right)$	建立在信息论的基础上
TF-IWF	$w_{ij}=tf_{ij}\times\left(\log\left(\frac{\sum_{i=1}^{M}nt_i}{nt_i}\right)\right)^2$	在 TF-IDF 算法的基础上，用特征项频率倒数的对数值 IWF 代替 IDF；并且用 IWF 的平方平衡权重值对于特征项频率的倚重

由于布尔权重（Boolean Weighting）计算方法无法体现特征项在文本中的作用程度，因而在实际应用中 0、1 值逐渐地被更精确的特征项的频率所代替。在绝对词频（Term Frequency，TF）方法中，无法体现低频特征项的区分能力，因为有些特征项频率虽然很高，但分类能力很弱（比如大多数目标共有的局部特征或背景特征），而有些特征项虽然频率较低，但分类能力却很强。

倒排文档频度（Inverse Document Frequency，IDF）法是文本分类中计算词

与文献相关权重的经典方法，其在信息检索中占有重要地位。该方法在实际使用中，常用公式 $L+\log((N-n_i)/n_i)$代替，其中，常数 L 为经验值，一般取为 1。IDF 方法的权重值随着包含某个特征的样本数量 n_i 的变化呈反向变化，在极端情况下，只在一个样本中出现的特征含有最高的 IDF 值。

本章使用的特征权重计算方法 TF-IDF，该方法的公式有多种表达形式，TFC 方法和 ITC 方法都是它的变种。实际应用中，有一种比较普遍的 TF-IDF 公式：

$$w_{ij}=\frac{tf_{ij}\times\log(N/n_i+0.01)}{\sqrt{\sum_{t_i\in T_j}[tf_{ij}\times\log(N/n_i+0.01)]^2}} \tag{6-26}$$

或

$$w_{ij}=\frac{(1+\log_2 tf_{ij})\times\log_2(N/n_i)}{\sqrt{\sum_{t_i\in T_j}[(1+\log_2 tf_{ij})\times\log_2(N/n_i)]^2}} \tag{6-27}$$

TF-IWF（Inverse Word Frequency）权重算法也是在 TF-IDF 算法的基础上提出的，其不同之处在于：

1）TF-IWF 算法中用特征频率倒数的对数值 IWF 代替 IDF；

2）TF-IWF 算法中采用 IWF 的平方来平衡权重值对于特征频率的倚重，不像 IDF 中采用的是一次方，给了特征频率太多的倚重。

此外，还有很多特征权重的计算方法，可以参阅文本分类的相关文献，这里不再一一列举。需要说明的是，权重计算方法与特征提取方法有着一定的关联，而很多文献引入的新的计算变量实质上都是考虑特征项在整个类中的分布问题。因此，需要进一步进行理论研究，获得更一般的有关特征权重确定的结论，而不是仅仅从不同的角度定义不同的计算公式。

6.6　实验结果与分析

1. 实验环境

（1）硬件环境

普通 DELL 台式计算机一台，基本配置为 P(R) D/3.4GHz/1.00G/160G/19in。

（2）软件环境

WindowsXP 操作系统，Visual Studio C++6.0 开发平台，OpenCV 函数库。

2. 实验数据来源

MSR 图像库包含海滩、瀑布、沙漠、山脉、建筑物、小汽车、花卉、水果、

飞鸟、蝴蝶等 22 类共计 3000 幅彩色图像，均存储为 JPEG 格式，大小为 352 × 231 像素至 352 × 530 像素不等。每一类图像数目为 42 ~ 289 幅，图像中目标的型号和姿态各异。

为了验证视觉单词库的性能，以及特征选择方法的效果，本章选用图像库中的 8 类图像分别求取在二分类问题上的实验结果。如图 6-6 所示，在进行汽车图像和建筑物图像分类时，挑选正负样本各 100 幅作为训练集样本，各 25 幅作为测试集样本，并挑选出 40 ~ 100 个正样本用以构造视觉单词库。训练集与测试集相互独立，即两者不含有同一幅图像。



图 6-6　训练集的正负样本示例

3. 分类器选用

目标识别系统中分类器的作用是：根据特征提取器得到的特征向量来给一个被测对象赋一个类别标记[52]。分类器的设计方法可以分为生成（Generative）方法和判别（Discriminative）方法两类。生成方法是根据类别出现的先验概率和条件概率来估计目标的类别概率，它将分类器设计问题转化为了概率密度估计问题，其代表是朴素贝叶斯分类器（Naive Bayesian Classifier，NBC）；而在判别方法中，将每个目标视为整个特征空间中的一个点，认为不同的类别是特征空间中不同的区域或子空间，需要找到一条决策边界把属于不同类别的点分开，其中最具代表性的是支持向量机（Support Vector Machine，SVM）和神经网络（Neural Network，NNet）。本章实验主要选用了朴素贝叶斯和支持向量机两种分类器，关于它们的基本原理和具体实现方法，在本书的 3.5 节已有介绍，此处不再赘述。

实验 1：视觉单词库构造方法对比分析

为了验证利用聚类算法构造视觉单词这一途径的有效性，本章将 RNN 凝聚聚类算法与划分方法中的 k-平均值和 k-中心点聚类应用于同一样本集，并比较最终的分类效果。该实验从 60 幅图片（小汽车图像）中共提取出 19127 个局部

特征（用高斯差分算子检测，并用 SIFT 描述子描述为 128 维的模式向量）用以构造视觉单词库，利用支持向量机分类器（线性核函数）对小汽车图像和建筑物图像进行分类测试，得到单词库规模为 200—1800 之间的正确率，该评估指标是在相等错误率（EER）下的分类效果，对图像进行向量空间模型表示时用的是词频权重。

如图 6-7 所示，由于 RNN 凝聚聚类算法得到的簇相对紧致，总体来说要比划分方法中的两种聚类算法性能好。关于视觉单词库的规模，在 200—800 之间随着视觉单词数量的增加分类效果得到了明显的改善，在 800 以上凝聚聚类算法相对稳定，*k*- 平均值和 *k*- 中心点方法则会出现波动，这是因为划分方法经常以局部最优结束。

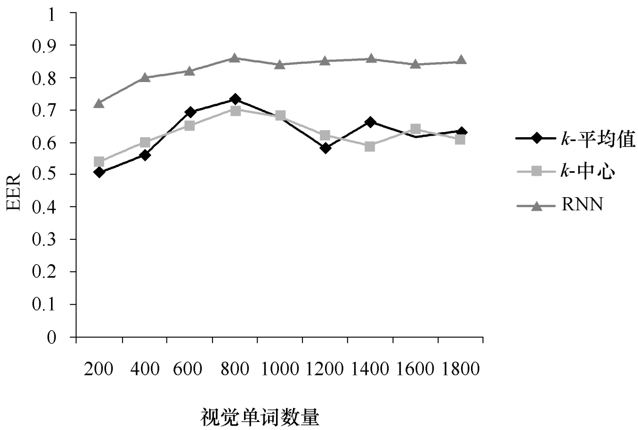


图 6-7　不同视觉单词库构造方法的性能

实验 2：图像数量和分类方法的效果分析

图像目标分类效果在受到视觉单词数量影响的同时，也和生成视觉单词库所用的图像数量有关。本实验从 40 幅图像提取出的局部特征是 10411 个，70 幅图像得到局部特征 21577 个，而 100 幅图像的局部特征数目达到 39350 个。在一定规模之内，从姿态各异的图像目标中提取越多的局部特征，构造出的视觉单词库内容就更为丰富，并且相应的视觉单词（原型特征）对分类来说更有区分性。

如图 6-8 所示，实验采用视觉单词的数量为 800 个，构造视觉单词库的图像为 30 ~ 100 个的时候，用朴素贝叶斯算法和支持向量机分别进行分类的效果。可以看出，在达到 60 幅图像的规模之后，图像的增加不再带来分类效果的明显改善；该实验的结果也简单证实了支持向量机在模式分类中的优

越性能。

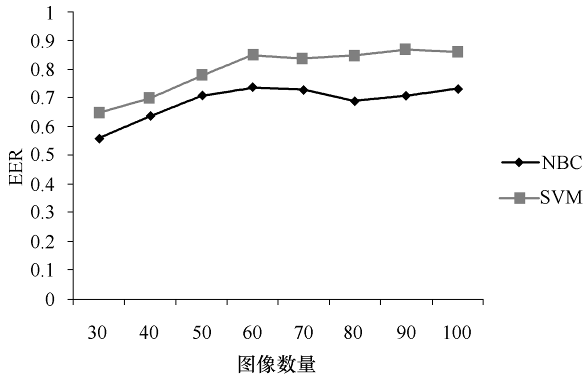


图 6-8　不同样本数量和分类方法的效果

实验 3：特征权重对分类效果的影响

采用不同的特征权重类型对分类的最终效果会有较大的影响，本章节将对布尔、绝对词频（TF）和 TF-IDF 三种特征权重计算方法进行实验对比。实验采用支持向量机（线性核函数）对 8 种图像目标分别进行二分类，求取每次分类的查准率和查全率。由于样本在所有类别中分布均匀，计算出的宏平均查准率和查全率等于微平均查准率和查全率。如图 6-9 的 RPC 曲线所示，该实验中布尔权重效果较差，而 TF 和 TF-IDF 权重效果相差不大。

由于用 0、1 来代表该视觉单词是否在图像目标中出现，布尔权重无法体现视觉单词在目标中的作用程度，因而分类效果显然不如更精确的 TF 方法。这从理论上讲，TF-IDF 作为一种相对词频权重，应当比 TF 的性能好，因为 TF 虽然体现了视觉单词的频率，但无法体现低频视觉单词的区分能力——有些视觉单词频率虽然很高，但分类能力很弱（比如大多数目标共有的特征或背景特征），有些视觉单词虽然频率较低，但分类能力却很强。但是从实验结果可以看出，TF-IDF 的效果不够理想，这一方面是因为图像目标分类中训练集的数目并不够大，本章在一次实验中训练样本只有 200 幅图像；另一方面很有可能是图像目标的向量空间模型表示维度较低，本章实验采用 800 维向量，这远远低于文本分类中所用的模式向量的维度。

实验 4：特征选择对分类效果的改善

特征选择在降低模式向量维数的同时保留了对分类有用的特征，本章节将通过图像频率（IF）、χ^2 统计量（CHI）方法、信息增益（IG）法和互信息（MI）法对图像特征进行筛选并进行分类效果对比。实验采用支持向量机（线性

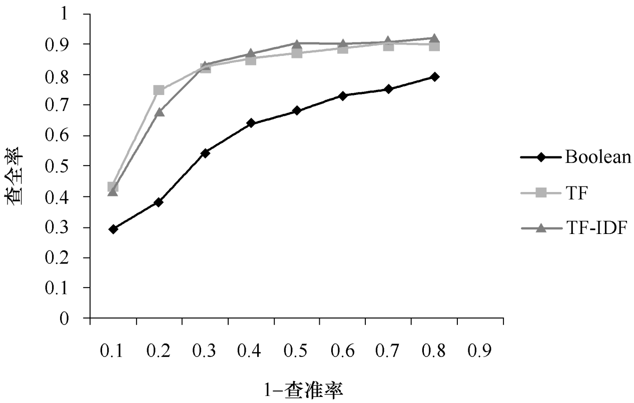


图 6-9　采用不同特征权重的分类效果

核函数）对图像目标进行二分类，在目标表示时用绝对词频计算特征权重，通过特征选择将视觉单词的数量从 800 减少至 450，步长为 50，测试每种方法在相等错误率（EER）下的分类正确率。如图 6-10 所示，在将特征维数降到 600 ~ 700 时大多数方法的效果最好，而总体看来基于互信息的特征选择方法性能较好。基于图像频率的特征选择方法最为简单易行，但该方法直接去除低频特征对分类效果产生不利影响。

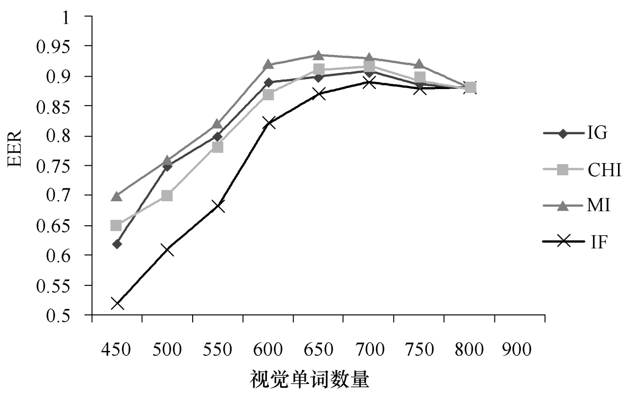


图 6-10　特征选择后的分类性能

实验 5：与相关文献的分类性能对比

近几年，国内外许多学者都在广泛关注利用局部特征进行图像目标识别这一研究方向。为了更为直观地比较 Weber、Opelt 等人提出的图像目标分类算法（参见参考文献［50］，［127］，［195］-［199］）与本书提出的算法的性能差异，进行如下对比实验。为了相关算法保持一致，实验所用的摩托车和小汽车（背

面视图）两类图像都选自 Caltech 图像库，算法的正确率是在相等错误率（EER）时计算所得的。从表 6-3 可以看出，与其他算法的最佳效果相比，本书算法的性能指标稍逊于 Zhang 提出的方法，总体看来正确率还是比较高的，可以说明本书算法的可行性。

表 6-3 相关文献算法与本书算法对比

Data Set	Motorbikes	Cars Rear
Weber（2000）	88.0%	—
Fergus（2003）	93.3%	90.3%
Opelt（2004）	92.2%	—
Thureson（2004）	93.2%	—
Deselaers（2005）	—	98.9%
Zhang（2007）	98.5%	98.3%
Leibe（2008）	94.0%	93.9%
Our algorithm	95.4%	96.8%

6.7 本章小结

由于局部特征性能优越，其携带的局部信息可以对图像的内容进行多语义层次的描述，本章尝试将局部特征应用于图像目标的分类识别。为此，作者对国内外大量相关科研成果进行了深入了解，并通过 RNN 凝聚聚类算法进行视觉单词库的构造。在此基础之上，充分借鉴了文本分类领域的向量空间模型进行目标表示，并结合信息论的相关技术进行特征优化，从而提出了一种基于局部特征的目标分类方法。在标准图像库上的实验结果证明了该方法的有效性和鲁棒性。

但是，本章提出的分类方法，在训练和识别过程中仅仅考虑将整幅图像或已分割好的区域作为目标，没有在图像中实现目标的自动检测与分割。这在很大程度上受限于向量空间模型不考虑特征项之间的空间关系的特点，从而造成了在视觉单词库的构造过程中，特征位置信息的缺失。下一步将考虑充分利用局部特征之间的空间关系，进行目标检测与分割的相关技术研究。

第 7 章　基于角点特征与视面模型的目标识别

如果一个理论本身具有持久性，那么最初给它带来很大威胁的那些反复辩难随着时间的推移只会有助于磨平它的粗糙之处，而如果有不抱偏见的、有见地的、真正平实的人士从事这一工作，甚至也可以使它在短时期内臻于所要求的精致优美。

——伊曼努尔·康德（1724—1804）

7.1　引言

视点不同造成目标的表象差异是图像目标识别领域的一个难点。在同一个场景中，视点的变化往往使得物体所呈现的表象有所不同，比如物体的大小比例、几何形状、物体的不同侧面等，这些都需要进行复杂的图像处理。对于视点远近变化造成的物体大小不同，要求识别系统具有某种尺度不变性，虽然通过尺度空间技术可以部分解决这个问题，但是，如何让计算机自动确定相应的尺度来识别物体，还需要相关科研人员的进一步研究与完善。

而由于观察的角度发生变化，同一物体的不同侧面呈现的特征往往大不相同，甚至产生了自身遮挡的问题（物体的某个部分遮挡了该物体的其他部分）。从物体自身的角度来说，也就是目标的不同姿态造成了模型库的模型数量激增，从而让目标识别的时间代价和空间代价变得异常昂贵。近年来，利用三维模型建立视面图的方法取得了一些成功。这种方法先是建立以三维目标为中心且与视点无关的 3D 模型，然后对视点进行限制并对目标进行平行投影得到二维视面模型，将目标可见表面相同的投影合并得到一个视区。针对不同视区可以提取目标在不同姿态下的特征，可以较好地解决目标姿态变化造成的目标难以识别的问题。

贝德曼[200]认为，当人们看物体时，会将其分割为一些简单的几何成分，称为几何元素（Geons）。他提出一共有 36 种这样的基本成分，并认为，有了这些基本的单元系列，我们就可以构建众多寻常物体的心理表征。他在物体知觉和言语知觉间进行了一番类比：利用英语的 44 个音素（Phonemes）或声音的基本单位，我们可以表现出英语中所有可能出现的单词（数量可达几十万）。同理，贝德曼认为运用基本几何元素也可以表现出成千上万的、立即就可以辨认的一般物体。

贝德曼还指出，当人们看见如图 7-1 所示的不完整的图画时，如果整个图中包括了物体的各个顶点——即这些片段还可以辨认出基本的几何元素的话，如图 7-1 中间一栏所示那样，那么人们还是能够确定所看到的是什么物体。但当这些顶点被删除以后（图 7-1 最右边的一栏），知觉者辨认基本几何元素的能力会受到影响，从而大大降低（几乎消减至零）正确辨认物体的可能性。

角点是一种图像的局部形状特征，只包含图像中大约 0.05% 的像素点，在没有丢失图像数据信息的条件下，最小化要处理的数据量。而且它具有旋转、平移、缩放不变性，几乎不受光照条件的影响，有很强的实用价值，已经被广泛应用于图像融合和图像拼接中，并取得了一系列成果。作为狭义特征点，角点不仅仅具有位置（Position）信息，还具有如下的其他信息：

1）夹角（Subtended Angle）：构成角点的二边界的夹角；

2）方向（Orientation）：角点夹角的角平分线方向；

3）边界形状（Edge Shape）：构成角点的边界是弧形还是直线形；

4）锐化度（Sharpness）：衡量边界在角点处的非连续性程度；

5）对比度（Contrast）：角点灰度与背景灰度的差值；

6）交点类型（Junction Type）：可分为 V 型、Y 型、T 型、K 型、X 型等。

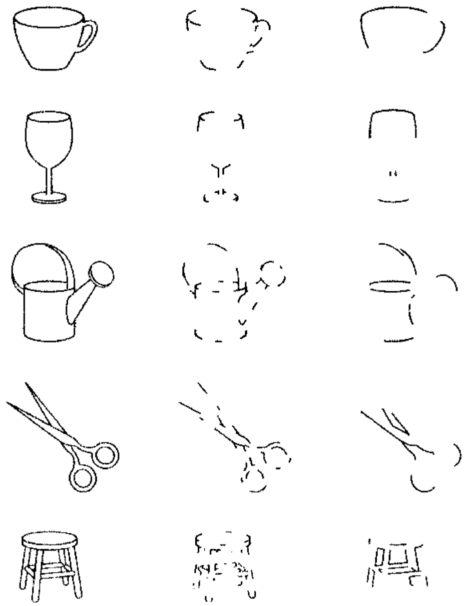


图 7-1　目标的几何元素（来源：贝德曼，1987 年）

虽然角点具有如此丰富的特征信息，但是在实际应用中人们往往只用到了角点的位置信息，抛弃了角点的其他特征。显然仅仅采用位置特征描述角点是极不充分的，这种不充分描述导致了角点在图像匹配中的局限性和容易产生误匹配。如果对角点进行更为丰富的描述，不仅可以加速匹配过程中的收敛过程和防止误匹配的发生，也会对在描述的过程中进行选择和优化，从而实现目标分类。

近年来，国内外一些学者都致力于利用角点的丰富信息构造合适的特征向量，初步应用于目标识别和图像分类。Baerveldt 等人[201]针对移动机器人的需要，采用角点作为识别物体的局部特征，设计了一个物体识别和定位系统。Dinesh 等人[202]通过角点及其三角空间关系（TSR）识别局部遮挡的物体。周振环[203]利用角点构造目标的多维距离特征向量，并应用于对飞机的识别。王鹏伟等人[204]提出了一种基于角点特征和自适应核聚类的目标识别方法，对遥感图像中的多个目标进行识别。

但上述方法只用到了角点的位置信息，抛弃了角点的其他信息，仅仅利用了角点之间的局部约束，描述方法过于简单，在实际应用中局限性非常大。本书将 Hausdorff 距离用于度量模板和目标的相似性，增强了目标匹配的抗噪声和抗遮挡能力，同时减少了识别的时间代价。同时充分利用角点间的全局约束和

局部约束得到类型可分离程度较高的特征向量，并根据目标的角点空间关系进行特征的选择和优化，结合 BP 网络强大的学习和泛化能力，实现不同姿态下的目标识别。

7.2 三维物体的视面模型表示

视面图（Aspect Graph）表示是一种用多个二维投影描述三维物体的方法。视面图方法的思想最早是由 Koenderink 等人[205]提出的，其核心概念是视面，它指的是一个物体在拓扑关系上等价的所有投影的代表性表示。对应于不同的观察空间，产生投影的方法分为两种类型：一种是把所有可能的视点定义在以目标为中心的单位球上，从球面上的一个点定义一个对中心点（目标）的观察方向矢量，用以产生目标的正交投影视图。另一种方法则考虑三维空间中的所有点，目标视图由视点的透视投影得到。

不论是哪一种投影方法，都可以通过偶然视点（Accidental Viewpoints）形成的边界将视点空间划分为一般视点（General Viewpoints）区域。对于一般视点，观察方向的变动并不会引起物体的视图变化（至少拓扑结构不会变化），所以也可称为稳定视点；而所谓偶然视点则相反，在这些视点上改变观察方向将得到不同结构的视图。从一般视点和偶然视点所得到的视图分别称为一般视图和偶然视图，从一般视点区域经过偶然视点边界进入另一个一般视点区域，称为一个视觉事件（Visual Event）。

视面图是一个图结构（图结构的概念，参见附录 B.1），其中每个节点代表目标的一个一般视图，每个弧表示两个相邻的一般视图之间的偶然视图或视觉事件。视面图被普遍认为是计算机视觉中一种很有潜力的表示方法，也已经研究出很多自动计算方法，用来得到多面体、曲面形体甚至具有任意连接的物体的视面图。但是迄今为止，对视面图方法的研究大都停留在理论阶段，其主要原因在于：视面图的数量可能会很大，因为对视面的检索代价太大；拓扑结构可以用数学的语言定义，但是却无法可靠地从图像中恢复。

国防科学技术大学的席学强[42]和陈晓飞[43]等人都在视面图的基础上对建立一般三维目标识别模型的方法进行了探索。他们采用了以三维目标为中心的且与视点无关的 3D 模型，通过对视点进行限制，并对目标进行平行投影得到二维视面模型。通过假定条件进行相应的简化，可以将目标视点范围限制为一个圆，称之为观察圆（View Circle）。以视点到目标质心的方向近似作为相机在该视点

处的光轴方向，投影平面与此方向垂直，用一组给定的图像平面上的正轴测投影（平行投影）来表示目标的二维视面。将目标可见表面相同的投影合并得到一个视区（View Region），视区就是由具有不会引起物体的视图变化的一般视点（稳定视点）构成的、被偶然视点包围的区域，视区所包含视面的数目为视区的长度。如此一来，就可以针对不同视区提取目标在不同姿态下的特征，在此基础上解决目标姿态变化造成的目标难以识别的问题。

图 7-2 所示为采用基于分裂-合并的层次聚类方法得到的三维目标的二维视区模型，图中右边的二维图像为目标的视区的原型视面。按照目标的复杂程度不同，可以将其用 6 ~ 10 个视区来表示。通过将相似的视面合并为视区并用原型视面来表示，减少视面的数量，提高了检索和识别的效率。

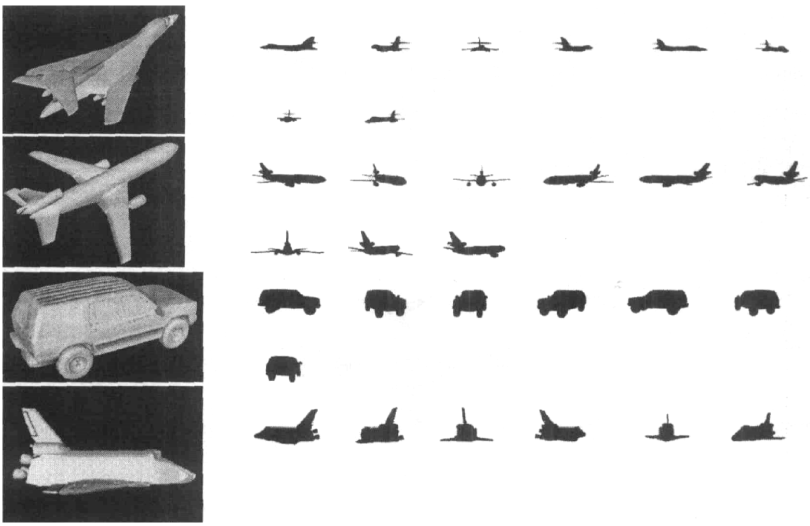
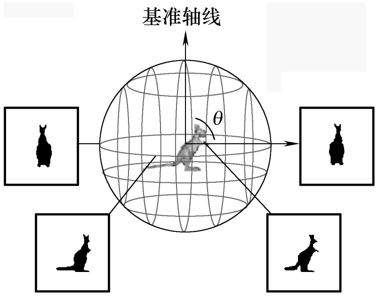
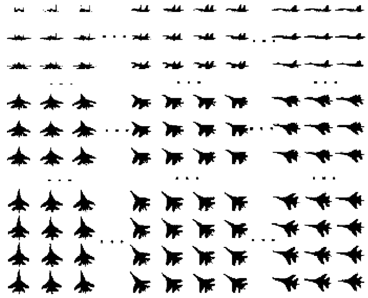


图 7-2　限制视点的三维目标的视区模型（来源：陈晓飞，2004 年）

本章也采用了类似的简化建模方法得到的实验所需的二维视面模型：首先建立以三维目标为中心的且与视点无关的 3D 模型，然后对视点进行限制并对目标进行平行投影得到二维视面模型。如图 7-3a 所示，以单位球上目标正上方的视点为基准视点，从该视点产生的正交投影视图为基准图像，基准轴线穿过基准视点和目标中心点。这样一来，每一个视点和目标中心的连线与基准轴线呈夹角 θ（锐角或直角），定义该夹角 θ 为视角。视角相同的视点同在一个观察圆上，目标姿态的变化程度随着视角的增大而加剧，这样就产生了三维物体的二维视面图。图 7-3a 中用作示例的观察圆（用红色表示的赤道线）的视角为 90°，图 7-3b 所示为 Su27 飞机模型在不同视点的投影图。



a)



b)

图 7-3　3D 模型的 2D 投影图表示

a）在单位球上进行投影　b）Su27 在不同视点下的投影图

7.3　基于角点特征的目标匹配

对目标最简单的描述方式就是视其角点为一个点集，它含有角点的数目和角点的位置信息。国内学者周振环[203]提出了通过角点的多维距离特征向量进行目标识别方法，多维距离向量就是通过计算每个角点到其余角点的距离得到的一组距离数值。这种方法比较简单，速度较快，但只具备平移和旋转不变性，不具有尺度不变性。王向军等人[206]进一步提出了通过特征角点构造特征描述的方法，特征角点即目标图像中最具有代表性、能简洁反映目标特征的角点。该方法虽然解决了尺度变化下的识别问题，但是仅限于对飞机的识别，灵活性不强，也不具备对局部遮挡目标的识别能力。

可见，良好的特征要具有区别性、可靠性、独立性和数目小这四个特点[168]，这也就意味着我们需要对角点点集做进一步处理，使其不受待识别目标的大小、位置、方位的影响，并适用于大多数目标种类。

7.3.1　利用基准角点进行目标匹配

对多维距离向量这种最具代表性的角点描述方法进行深入分析后，不难发现，数量繁多的角点不仅增加了运算时间，同时也给多帧图像中的目标匹配带来了困难，尤其是有噪声干扰和在目标姿态变化的情况之下。为了达到更加高效、灵活的识别效果，本书希望能够选择出目标图像上最具有代表性的角点，根据这些基准角点的位置信息测量物体，并通过测量值来识别目标。

以军事目标飞机为例，测量一个飞机可以利用的最显著的信息就是机头部分、两翼部分和机尾部分的角点，以及它们和飞机重心的位置关系。飞机重心

的计算公式如下：

$$\begin{cases} x_G = \sum_{x,y\in R} xI(x,y) / [\sum_{x,y\in R} I(x,y)] \\ y_G = \sum_{x,y\in R} yI(x,y) / [\sum_{x,y\in R} I(x,y)] \end{cases} \tag{7-1}$$

式中，(x_G, y_G)是重心点 G 的位置坐标，$I(x, y)$ 是像素点（x, y）的灰度值。我们可以通过计算每个角点的相对重心的重心矩，并选择重心矩最大的角点作为第一基准角点 P_1（x_{p1}, y_{p1}）：

$$M_{x_{ci},y_{ci}} = \sum_{ci\in C} [(x_{ci} - x_G)^2 + (y_{ci} - y_G)^2] I(x_{ci}, y_{ci}) \tag{7-2}$$

$$M_{x_{p1},y_{p1}} \geqslant M_{x_{ci},y_{ci}}, p_1, c_i \in C \tag{7-3}$$

式中，C 为角点点集，其包含的角点数目为 n。

通过重心点 G 和 P_1 可以得到飞机的一条基准轴线，显然，飞机两翼上的角点都位于这条轴线的两侧。于是这条轴线就把机翼角点点集划分为两个子集：

$$\begin{cases} (x_{p1} - x_G)(y_{ci} - y_G) - (y_{p1} - y_G)(x_{ci} - x_G) > 0 \\ (x_{p1} - x_G)(y_{ci} - y_G) - (y_{p1} - y_G)(x_{ci} - x_G) < 0 \end{cases}, i = 1, 2, \cdots, n-1 \tag{7-4}$$

而两翼上的基准角点就可以定义为距离这条轴线最远的点——第二基准角点 $P_2(x_{p2}, y_{p2})$ 和第三基准角点 $P_3(x_{p3}, y_{p3})$。角点与基准轴线的距离可以由以下公式计算得到：

$$D_{ci} = \frac{|(y_{p1} - y_G)(x_{ci} - x_G) + (x_{p1} - x_G)(y_G - y_{ci})|}{\sqrt{(y_{ci} - y_G)^2 + (x_{ci} - x_G)^2}}, i = 1, 2, \cdots, n-1 \tag{7-5}$$

第四基准角点 $P_4(x_{p4}, y_{p4})$ 被定义在机尾，它和第一基准角点 P_1 分别位于基准轴线的两端，也就是重心点 G 的两侧。如果 P_1 满足以下条件：

$$-\frac{x_{p1} - x_G}{y_{p1} - y_G} x_{p1} - y_{p1} + \left(y_G + \frac{x_{p1} - x_G}{y_{p1} - y_G} x_G\right) > 0 \tag{7-6}$$

则 P_4将满足

$$-\frac{x_{p1} - x_G}{y_{p1} - y_G} x_{p4} - y_{p4} + \left(y_G + \frac{x_{p1} - x_G}{y_{p1} - y_G} x_G\right) \leqslant 0 \tag{7-7}$$

否则，必然有

$$-\frac{x_{p1} - x_G}{y_{p1} - y_G} x_{p4} - y_{p4} + \left(y_G + \frac{x_{p1} - x_G}{y_{p1} - y_G} x_G\right) > 0 \tag{7-8}$$

在目标识别过程中，特征空间优化的目的是用最少的描述获得目标形状上最“本质”的特征。通过每个基准角点到重心点 G 的距离，可以定义出一个区

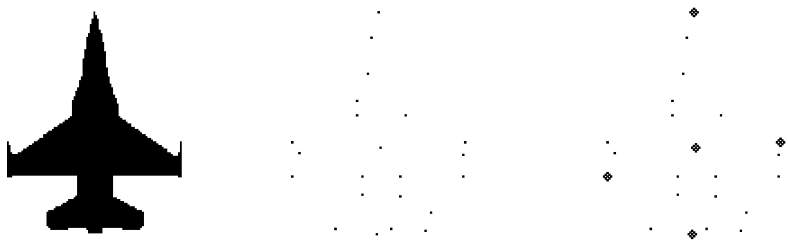
分度较高的描述子：

$$S = S_1 \times S_2 \tag{7-9}$$

$$S_1 = \frac{|L_{G2}|}{|L_{G3}|} = \frac{\sqrt{(x_{p2} - x_G)^2 + (y_{p2} - y_G)^2}}{\sqrt{(x_{p3} - x_G)^2 + (y_{p3} - y_G)^2}} \tag{7-10}$$

$$S_2 = \frac{|L_{G4}|}{|L_{G1}|} = \frac{\sqrt{(x_{p4} - x_G)^2 + (y_{p4} - y_G)^2}}{\sqrt{(x_{p1} - x_G)^2 + (y_{p1} - y_G)^2}} \tag{7-11}$$

如图 7-4 所示，用 SUSAN 算法检测出了 F16 战斗机图像的所有角点，并按照上述方法获得了四个基准角点，从而构造出了特征描述子 S。



a)　　b)　　c)

图 7-4　飞机图像的基准角点提取

a）F16 战斗机　b）角点（SUSAN）　c）4 个基准角点

7.3.2　基于主分量与 Hausdorff 距离的匹配算法

霍特林（Hotelling）[207]提出了一个可以去掉一个随机向量中各元素间相关性的线性变换，并把它称作“主分量法”。此后，卡胡南（Karhunen）和列夫（Loeve）提出了一种针对连续信号的类似的变换。这种方法派生出了一种离散图像变换的方法。

我们根据角点的坐标可以生成二维向量，可以把这些二维向量当成原理中的随机向量 $X = (a, b)^T$处理，其中 a 和 b 是角点关于 x_1 轴和 x_2 轴的坐标值。总体的均值向量（边界点）可以通过 K 个样本向量（角点）来估计：

$$m_X = E\{X\} \approx \frac{1}{K}\sum_{k=1}^{K} X_k \tag{7-12}$$

总体向量的协方差矩阵可以以如下方式用样本近似得到：

$$C_X = E\{(X - m_X)(X - m_X)^T\} \approx \frac{1}{K}\sum_{k=1}^{K} X_k X_k{}^T - m_X m_X{}^T \tag{7-13}$$

因为 C_X是实对称的，找到一组 n 个标准正交特征向量总是可能的。令 e_i和 λ_i，$i = 1, 2, \cdots, n$ 为特征向量和对应的 C_X特征值，以降序排布，使 $\lambda_j \geqslant \lambda_{j+1}$，

$j=1, 2, \cdots, n-1$。令 A 为一个由 C_X 的特征向量组成其行元素的矩阵，并进行排序，使 A 的第一行为对应最大特征值的特征向量，而最后一行为对应最小特征值的特征向量。假设把 A 作为将 X 的向量映射到用 y 代表的向量的变换矩阵，就得到了霍特林变换的表达式：

$$y = A(X - m_X) \tag{7-14}$$

使用式（7-14）的实际结果是需要设置一个新的坐标系统，这个坐标系统以角点总体的质心（均值向量的坐标）为原点，以 C_X 的特征向量所指方向为轴的方向，如图 7-5b 所示。这个坐标系统清晰地显示出式（7-14）所进行的变换是一种旋转变换，这种变换使用特征向量将数据排列起来，如图 7-5c 所示。实际上，这种排列正好是数据去相关的机理。另外，由于特征值沿着 C_y 的主对角线排列，λ_i 是沿着特征向量 e_i 的分量 y_i 的方差，这两个特征向量是正交的。由于这个明显的原因，y 轴被称为本征轴[8]。

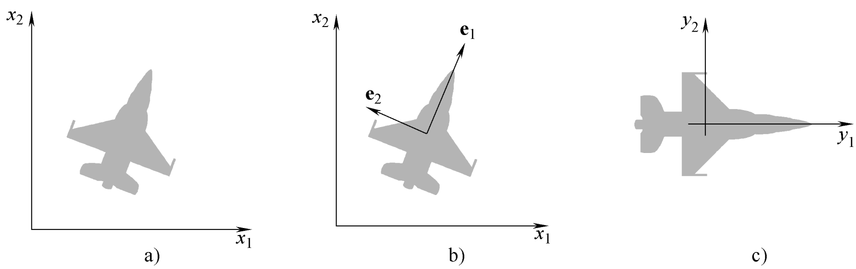


图 7-5　用主分量法将目标沿着自身的本征轴对准

a）一个目标　b）特征向量　c）旋转目标

使用主特征向量排列角点的概念在图像描述中起着十分重要的作用。正如前面提到的，目标的描述对于大小变化、平移和旋转变化本应是尽可能独立的。使用目标的主轴校正的能力为消除旋转变化的影响提供了一种可靠手段。特征值是沿着本征轴的方差，并可用于尺寸的归一化。平移带来的影响可以通过将角点的均值设定为中心来解决。

Huttenlocher 等人[208]提出的 Hausdorff 距离是用来描述两组点集之间相似程度的一种度量，是集合与集合之间距离的一种定义形式。它与许多其他匹配算法不一样，它并不要求目标与模板的简单一致，而是可以针对部分匹配作出良好的反应，因此它本身就具有一定的抗遮挡能力。对有限点集 $A=\{a_1, a_2, \cdots, a_p\}$ 和 $B=\{b_1, b_2, \cdots, b_p\}$，$A$，$B$ 之间的 Hausdorff 距离定义如下：

$$H(A,B) = \max[h(A,B), h(B,A)] \tag{7-15}$$

$$h(A,B) = \max_{a\in A} \cdot \min_{b\in B} \| a-b \| \tag{7-16}$$

$$h(B,A)=\max_{b\in B}\cdot\min_{a\in A}\|b-a\| \tag{7-17}$$

式中，$H(A, B)$ 是 $h(A, B)$、$h(B, A)$ 中较大的那一个，称为 A、B 之间的 Hausdorff 距离；$h(A, B)$ 称为点集 A 到 B 的有向 Hausdorff 距离，即点集 A 中的每个点 a_i 到 B 集中与其距离最近的点 b_j 之间的距离 $|a_i-b_j\|$ 进行排序，取这样的距离中的最大值作为 $h(A, B)$ 的值，同理可得 $h(B, A)$；$|^*\|$ 表示某种距离范数，如欧氏距离。如图 7-6 所示，Hausdorff 距离表征了两个点集之间的最大不相似程度。

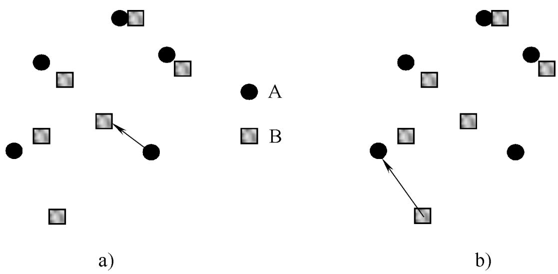


图 7-6　Hausdorff 距离示意图

a）表示点集 A 到 B 的有向 Hausdorff 距离　b）表示点集 B 到 A 的有向 Hausdorff 距离

在本书的应用中，为了降低噪声的影响，我们使用部分 Hausdorff 距离，其定义如下：

$$H_{LK}(A,B)=\max[h_L(A,B),h_K(B,A)] \tag{7-18}$$

$$h_L(A,B)=L^{th}_{a\in A}\cdot\min_{b\in B}\|a-b\| \tag{7-19}$$

$$h_K(B,A)=K^{th}_{b\in B}\cdot\min_{a\in A}\|b-a\| \tag{7-20}$$

式中，H_{LK}（A，B）仍是 h_L（A，B）和 h_K（B，A）中较大的一个。h_L（A，B）虽然还是按照 $|a_i-b_j\|$（即 A 中的每个点 a_i 到 B 中与其距离最近的点 b_j 之间的距离）进行排序，但不是像 h（A，B）那样取全局最大值，而是取第 L 个值（$1\leqslant L\leqslant q$，q 为 A 集中点的数目），h_K（B，A）同理可得。

通过角点检测算法可以得到待匹配目标和原型的两组特征点集，则目标匹配问题就转化为特征点匹配问题。因为 Hausdorff 距离的适用形式限制在有限点集内，所以非常适合度量特征点集的相似性。而角点点集经过主分量法处理后，消除了其对尺寸、位置、方位的依赖性，就可以作为 Hausdorff 距离的匹配元素，对这些元素的相似性进行度量并将此度量值作为目标与原型相似性的依据，如此一来，大大降低了算法的运算复杂度，并减少了噪声对识别效果的影响。

7.4　基于角点标记图的目标分类

统计模式识别把模式类看成是用某个随机向量实现的集合，是对模式的统计分类方法，又称决策理论识别方法。模式即描绘子的组合，模式类是一个拥有某些共同性质的模式簇，实践中的三种常用模式组合是向量（用于定量描述）、串和树（用于结构描述）。而模式向量是统计模式识别最为常用的表示形式，其元素是根据什么量的描绘子进行选择的，对于目标识别的最终效果有很大影响。

7.4.1　角点特征的优化技术

确定合适的特征空间是设计目标识别系统的一个十分关键的问题。如果所选用的特征空间能使同类物体分布具有紧致性，不同类别物体彼此分开，即各类样品能分布在该特征空间中彼此分隔开的区域内，这就为分类器设计提供良好的基础。反之，如果不同类别的样品在该特征空间中混杂在一起，再好的设计方法也无法提高分类器的准确性。对特征空间进行优化有两种基本方法，一种是特征选择，即对原特征空间进行删选，另一种就是特征的组合优化，即通过一种映射变换改造原特征空间。

通过对角点检测结果的仔细观察和分析，我们发现，过于密集的角点往往会局限于个例的细节变化，在训练分类器的时候容易产生过拟合现象。而且有一些点是图像获取或传输中产生的噪声，直接对目标的特征描述产生干扰。因此，为了减少分类器的训练复杂度、增强系统的鲁棒性，对特征空间进行适度的优化，是十分必要的。

对于直线投影法检测到的角点，按照在轮廓线上的顺序，如果一个角点与其前后两个角点的距离很近，且这三个相邻的角点和形心的距离相等或接近，则该角点所携带的信息与前后角点有冗余，可以删选掉。于是我们可以通过角点和质心的空间关系，计算出每个角点对于整个形状特征的重要程度，并据此对角点进行筛选：

$$w_i = \omega_1 \frac{|d_{i+1} - d_i| + |d_i - d_{i-1}|}{d_{\max}} + \omega_2 \frac{D_{i-1,i} + D_{i,i+1}}{D_{\max}} \tag{7-21}$$

式中，d_i表示第 i 个角点到形心的距离，$D_{i-1,i}$表示第 i-1 个角点和第 i 个角点（按照角点在轮廓线上的顺序）的距离，$d_{\max}$是角点到形心的最大距离，$D_{\max}$是相邻两个角点间的最大距离，ω_1和 ω_2是该项的权重。

由于 SUSAN 算法不依赖于目标分割得到的轮廓信息，所以无法通过跟踪轮廓来得到角点的顺序，并依此计算每个角点的权重，进行特征空间的优化。但

是，我们依然可以从角点负载信息量的角度考虑，将非常密集的角点群用一个点来代替，保留孤立、信息量巨大的角点，从而在优化特征空间的同时保持了目标的基本几何形状。

聚类分析是机器学习领域的一个重要研究方向，目前存在大量的聚类算法，算法的选择取决于数据的类型、聚类的目的和应用。本书选用凝聚的层次聚类方法[128]对原特征空间进行组合优化，以求出一组对分类识别更为有效的特征。这种自底向上的策略首先将每个角点作为一个簇，然后将相似度最大的原子簇合并，直至达到某个希望的簇的数目。簇间相似度是通过计算平均相似度（一个簇中所有对象和另一簇所有对象之间的相似度的平均）得到的：

$$sim(C_1,C_2) = \frac{1}{|C_1||C_2|}\sum_{p_1 \in C_1}\sum_{p_2 \in C_2} sim(p_1,p_2) \tag{7-22}$$

其中，相似度的度量采用的是欧氏距离。最终，用每个簇的重心（簇的所有角点的平均值）来代表整个簇。两种角点检测算法及其相应的特征空间优化方法的效果，如图 7-7 所示。

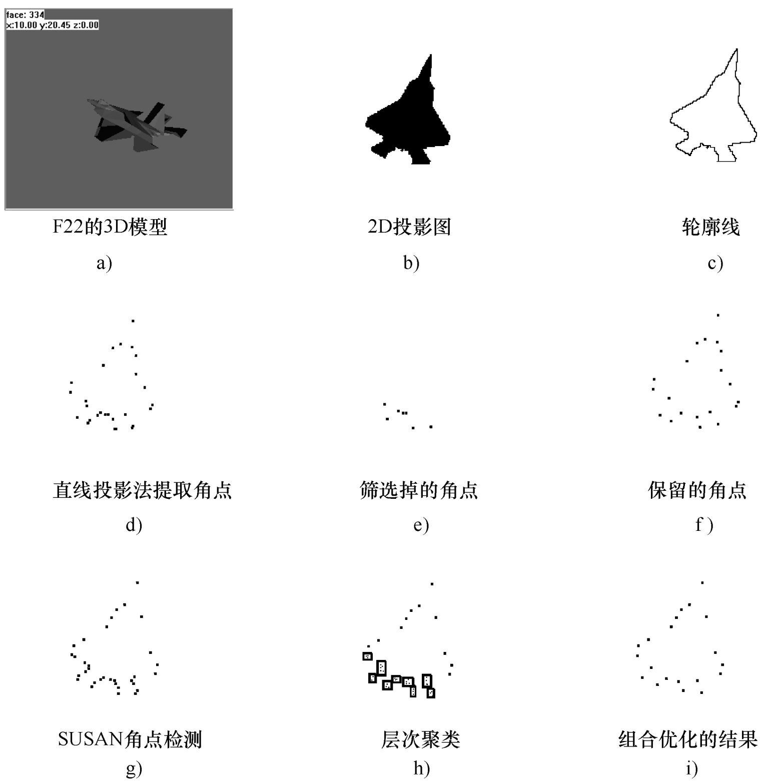


图 7-7　对角点特征空间进行优化设计

（d）~e）是进行特征选择，g）~i）是进行特征的组合优化）

7.4.2　角点标记图的生成方法

标记图是一种一维函数的边界表达方法，其典型的生成方法是将从质心到边界线的距离转化成一个角度函数，如图 7-8 所示。虽然其生成方法多种多样，但基本思想都是，假设一维函数表达会比原来的二维边界容易，因此使用一维函数简化边界的表达[8]。本书提出的角点标记图将标记图的基本思想应用于构造角点特征的过程中，并在保存目标基本信息的同时消除其对尺寸和旋转的依赖性，使得该特征具有平移不变形、比例不变性和旋转不变性。

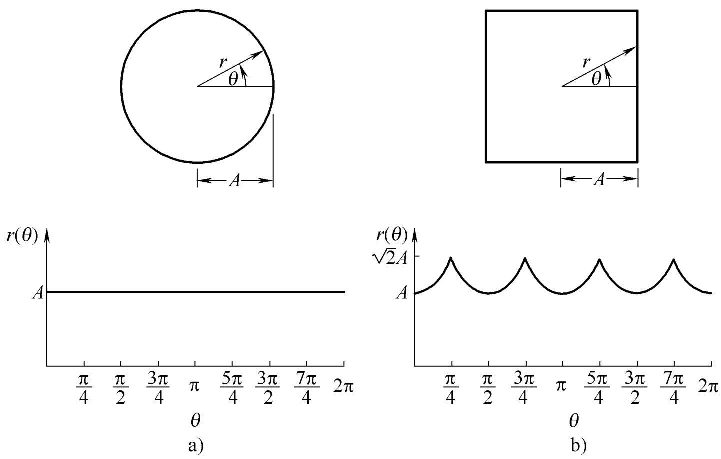


图 7-8　距离-角度的函数标记图

a）圆形的标记图　b）正方形的标记图

假设经过特征选择或特征的组合优化之后，最终得到了 N 个角点，则特征向量可以初步表示为下列形式：

$$\boldsymbol{x} = [x_1, x_2, \cdots x_n]^{\mathrm{T}} \tag{7-23}$$

这里，每个分量 x_i 代表第 i 个角点到质心的距离。

这种模式向量的生成方法依赖于旋转和比例缩放变换。需要寻找一种方法，选择相同的起点而忽略图形的方向，实现旋转变换的归一化。可以选择距离质心最远的点作为起点，如果这一点与我们关心的每个图形的旋转畸变无关，或者按照距离质心的远近对角点进行排序（角点在轮廓上的次序缺失的情况下）。由于图形尺寸变化会导致对应特征向量的分量值的变化，将这种结果进行归一化的一种方法就是，对所有分量值进行换算，以便向量的各个分量有相同的值域，比如［0，1］。这种方法的主要优点是简单易于实

现，当然它也有潜在的缺陷，即对所有分量的缩放仅依赖于两个值：最小值和最大值。如果图形是带有噪声的，这种依赖性就可能成为从对象到对象的误差来源。

7.5 实验结果与分析

1. 实验环境

（1）硬件环境

普通 DELL 台式计算机一台，基本配置为 P（R）D/3.4GHz/1.00G/160G/19in。

（2）软件环境

WindowsXP 操作系统，Visual Studio C + +6.0 开发平台，OpenCV 函数库。

2. 实验数据来源

普林斯顿大学三维模型库（Princeton Shape Benchmark）[209]经常被用来研究与 3D 模型相关算法的优劣。这个公用平台中提供了多达 1800 个三维模型作为待识别目标，如图 7-9 所示，可以对其进行不同视角的投影拍摄来建立一个符合自己实验要求的目标图库。

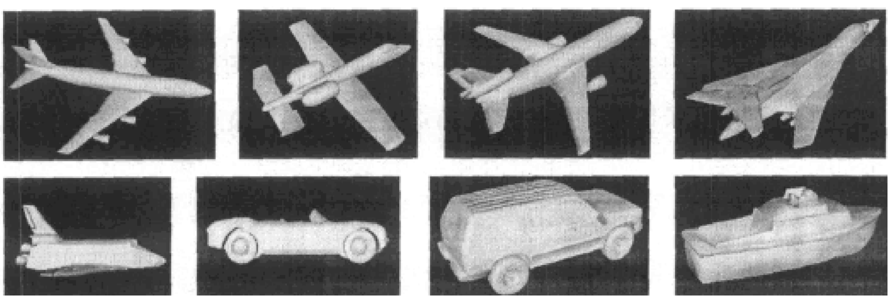


图 7-9 普林斯顿大学三维模型库示例

本书从普林斯顿大学三维模型库中挑选出六个飞机模型、一个 T60 坦克模型、一个小汽车模型用来衡量目标识别系统的性能，其中六个飞机模型的型号依次是 F16，F117，M1237，747G，F1，F2。图 7-10 就是采用 7.2 节提出的方法获得的基准视点下的 3D 模型 2D 投影图。

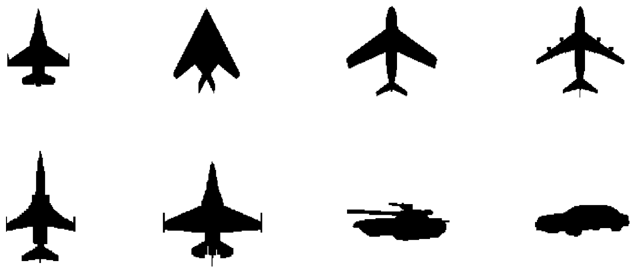


图 7-10　3D 模型的 2D 投影图表示

实验 1：利用基准角点进行目标匹配

由于基准角点的提出是针对飞机图像的，本实验选用了六个飞机模型在不同视点下的投影图：在每个模型的单位球上，从北纬 90°到北纬 60°，纬度每隔 5°、经度取固定值（与机身最长轴线垂直）进行一次投影，每个模型都得到 7 幅投影图，样本数量共计 6 ×7 =42。我们使用每个飞机模型在基准视点的投影图作为模板，并通过 F16 战斗机的所有样本对其进行相似性度量，表 7-1 是 F-16战斗机投影图（不同视点下）与其他飞机投影图（基准视点下）的特征描述子 S 的比例。1 表示样本与模板完全吻合，数值与 1 做差的绝对值越大则表示样本与模板越不匹配。

表 7-1　利用基准角点进行目标匹配的相似度结果

模型 / θ	F16	F1	F2	M1237	747G	F117
0°	1.000	0.678	0.945	0.763	0.782	0.675
5°	0.974	0.696	0.920	0.744	0.761	0.693
10°	0.944	0.719	0.892	0.721	0.738	0.715
15°	0.943	0.719	0.891	0.720	0.738	0.715
20°	0.942	0.721	0.889	0.719	0.736	0.717
25°	0.903	0.751	0.853	0.689	0.706	0.747
30°	0.930	0.729	0.879	0.710	0.727	0.726

由实验结果可以看出，基准角点作为目标图像中最具有代表性、能简洁反映目标特征的角点，抓住了目标形状上最“本质”的特征。通过每个基准角点到重心点 G 的距离，也可以定义出一个具有平移、旋转、尺度不变性描述子。但在进一步应用中会发现，坦克、车辆等目标的基准角点很难定义，或者说各个角点的作用不像飞机有如此显著的差别。这就导致了这种特征描述子只能在

特定目标（例如飞机）的识别中发挥作用。灵活性不强，也不具备对局部遮挡目标的识别能力。

实验 2：基于主分量与 Hausdorff 距离的目标匹配

本实验为了验证基于主分量与 Hausdorff 距离的目标匹配算法的通用性和鲁棒性，将八个三维模型在不同视点下的投影图：在每个模型的单位球上，从北纬 90°到北纬 60°，纬度每隔 5°、经度取固定值（与机身最长轴线垂直）进行一次投影，每个模型都得到 7 幅投影图，样本数量共计 8 ×7 =56。我们使用每个三维模型（如 F117 战斗机、汽车、坦克等）在基准视点的投影图作为模板，并通过 F16 战斗机的所有样本对其进行相似性度量。表 7-2 是 F16 战斗机投影图（不同视点下）与其他三维模型投影图（基准视点下）的 Hausdorff 距离，距离为 0 表示样本与模板完全吻合，距离越大则表示样本与模板越不匹配。

表 7-2　目标与各个模型的 Hausdorff 距离

模型 / θ	F16	F117	M1237	747G	F1	F2	T60	car
0°	0. 000	18. 385	48. 104	15. 232	14. 036	18. 788	68. 154	69. 584
5°	6. 325	18. 028	49. 578	16. 763	13. 416	18. 385	70. 093	71. 568
10°	10. 000	17. 804	50. 329	17. 720	15. 232	15. 811	71. 063	72. 560
15°	12. 166	18. 000	51. 088	18. 682	15. 264	18. 682	72. 035	73. 552
20°	15. 556	19. 000	52. 631	18. 385	17. 720	17. 464	73. 980	75. 538
25°	15. 556	19. 925	52. 631	21. 213	20. 396	18. 439	73. 980	75. 538
30°	17. 117	22. 023	52. 631	25. 179	24. 187	21. 024	73. 980	75. 538

实验结果证明，基于主分量与 Hausdorff 距离的目标匹配算法能够很好地识别出不同类别的目标，即非常明显地区分出飞机和车辆，即使视点发生了高达 30°的变化。但进一步对同类目标的识别效果相对逊色，比如，F16 在基准视点与其在 30°视点的投影图之间的距离显然要比 F16 与 F1 或 F2 在基准视点的投影图距离要大，这样就非常容易产生错误的匹配。可见，在视点发生变化的情况下，基于 Hausdorff 距离的目标匹配方法对于不同种类的刚性目标有着很好识别效果，而对同类目标的识别并不十分理想。

实验 3：基于角点标记图与 BP 网络的目标分类

近年来，傅里叶描绘子、标记图和不变矩特征被广泛应用于目标识别领域。相对传统的矩形度、圆形度等描述方法，这三种特征不仅对复杂形状有着更好

的逼真度，而且解决了平移、尺度和旋转不变性问题，基本上可以满足多数情况下形状匹配和目标分类的需求。但目标姿态变化导致的形状改变，是对传统的识别方法的新挑战，也对特征描述提出了更高的要求。

本实验将角点标记图与以上三种形状特征进行了实验对比，实验中使用相同的图像数据、同一种分类器——三层 BP 网络，特征向量的维数都为 20（除不变矩的描述通常为 7 维）。选用的飞机 3D 模型为 F16，F117，M1237，747G，在每个模型的单位球上，从北纬 90°到北纬 60°，纬度每隔 5°、经度每隔 2°进行一次投影，则在每个纬度下，共有 360/2 = 180 幅投影图，我们随机抽取其中的 120 幅作为训练集，余下的 60 幅为测试集。由于共有 4 个 3D 模型，每个模型在 7 个纬度下的进行投影，整个训练集和测试集样本数量分别为：120 × 7 × 4 = 3360，（180 − 120）× 7 × 4 = 1680。作为基准视点，模型的正上方北纬 90°对应于 $\theta = 0°$ 的观察圆，北纬 85°则对应与 $\theta = 5°$ 的观察圆，以此类推。此处采用整个训练集的一部分数据对分类器进行训练，只是在基准视点（$\theta = 0°$）的投影图，共 120 × 4 = 480 幅。测试时用的是全部的测试集数据，即 1680 幅投影图。

实验结果如图 7-11 所示，两种角点标记图在视点发生变化的时候，对目标的识别效果好于其他三种特征。虽然在视角变化较小的时候（$\theta = 5°$），标记图的错误分类率小于角点标记图，但是当视角变化逐渐增大的时候，角点标记图的稳定性和识别率的优势愈加明显。

实验 4：增加训练模式后的分类效果改进

通过下面的训练方法尝试进一步提高系统的识别能力：在使用基准视点（$\theta = 0°$）的投影图进行训练之后，用余下的训练集数据对系统进行重新训练，再

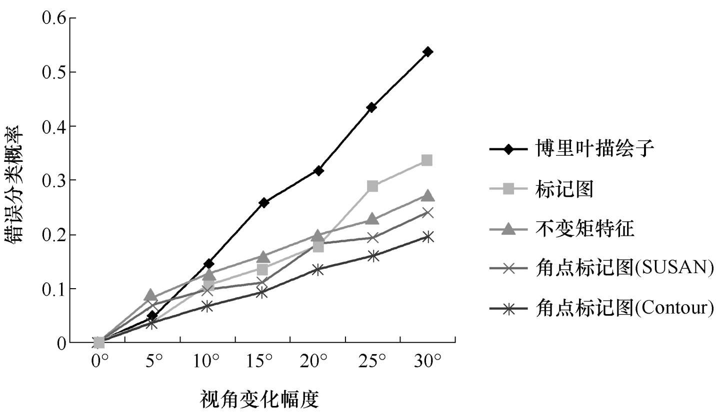


图 7-11　五种特征在视点变化下的识别效果

用新权值向量通过系统运行测试集样本来确立识别性能。图 7-12 给出了 SUSAN 角点特征通过持续这种再训练和令 $\theta=5°$，10°，15°，20°，25°和 30°后进行的再测试过程得到的结果。

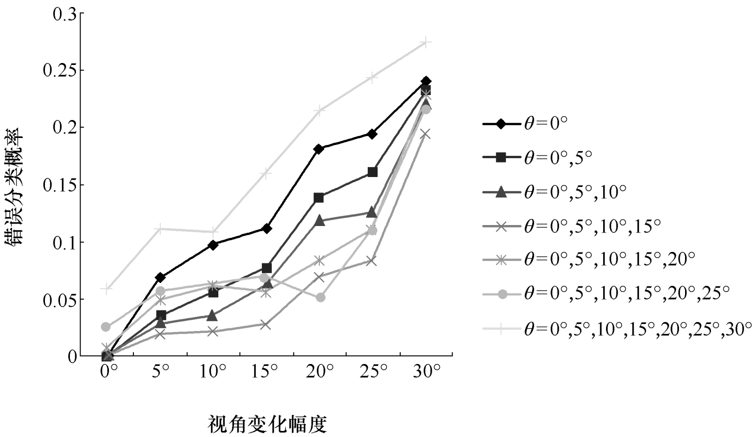


图 7-12　通过增加训练模式改进识别效果

正如所希望的，如果系统进行了适当的学习，由测试集检测出的错误分类概率会在增大时降低，故通过系统性的、视点变化幅度小量增加的训练来增强系统的分类能力是很重要的。但如果视点变化幅度超过了 15°，目标的形状变化剧烈甚至产生了自身的局部遮挡，这种情况下神经网络在学习过程中的收敛和稳定性就很不理想了。

7.6　本章小结

角点具有位置、夹角、方向、边界形状、锐化度、对比度、交点类型等丰富的信息，在图像匹配和图像拼接领域中已经取得了显著的成果。本章通过基准角点和重心的相对位置信息测量物体，定义出一个具有平移、旋转、尺度不变性描述子来识别目标，该方法有一定的健壮性，非常适合于特定目标（例如飞机）的识别；本章还结合主分量法和 Hausdorff 距离，提出了一种在视点变化下目标匹配识别方法，不仅增强了识别算法的抗噪声和抗遮挡能力，同时也减少了识别的时间代价。

为了有效利用众多样本进行分类识别，本章在对特征空间进行优化设计的基础上，提出了一种基于质心的角点特征构造方法——角点标记图，这种特征

简单有效地反映了目标的形状特性，并具有平移、旋转、尺度不变性以及对噪声的抗干扰能力。与其他三种形状特征进行实验结果对比，表明采用该特征的分类方法在视点发生变化时对目标的识别更为稳定、有效，且通过系统性的、视点变化幅度小量增加的训练可以进一步增强其目标识别的能力。

附录 A　图像处理的一些相关理论

A.1　数字图像的基本概念

“图像”一词在汉语中很难给出一个明确的定义，它在英文中有三个相关词汇——“picture”、“image”和“pattern”。英文词典一般是这样注释的，picture——画、图画、图像、图片、电影等；image——像、图像、景像、映像、影像、反射、映射等；pattern——模型、式样、样本、图案、花样、图、图形等。从这三个词的注释中大致可做如下区分，“picture”是指与照片等相似的用手工描绘的人物或景物，其中侧重于手工描绘的一类“画”；“image”是指用镜头等科技手段得到的视觉形象，一般来讲可定义为“以某一技术手段被再现于二维画面上的视觉信息”，通俗地说就是指那些用技术手段（包含计算机技术）把目标原封不动地一模一样地再现的图像；而“pattern”在拉丁语中指裁衣服的纸样，因此它主要指的是图案、曲线、图形。综上所述，我们说的图像应该是“image”，“Image Processing”处理的主要是照片、复印机、电视机、传真机、计算机显示的一类图像[104]。

“图像”和“图象”这两个名词易于混淆，在各种专业书籍里面也经常混用，如果一定要做辨识，我们可以简单地认为，“图象”一般用于表示数学领域中的图，如函数图象一类的东西。而对于“图像”和“图形”这两个概念，我们可以从以下几个方面进行区分：

1）存储方式的区别。图形存储的是画图的函数。图像存储的则是像素的位置信息和颜色信息以及灰度信息。

2）缩放的区别。图形在进行缩放时不会失真，可以适应不同的分辨率。图像放大时会失真，可以看到整个图像是由很多像素组合而成的。

3）处理方式的区别。对图形，我们可以旋转、扭曲、拉伸等。而对图像，我们一般会进行对比度增强、边缘检测等。

4）算法的区别。对图形，我们可以用几何算法来处理。对图像，我们可以用滤波、统计的算法。

5）其他。图形不是主观存在的，是我们根据客观事物而主观形成的。图像则是对客观事物的真实描述。

当用数学方法描述图像信息时，通常着重于考虑它的点的性质。例如一幅图像可以被看成是空间各个坐标点上强度的集合。它的最普遍的数学表达式为

$$I = f(x, y, z, \lambda, t) \tag{A-1}$$

式中，(x, y, z)是空间坐标，λ 是波长，t 是时间，I 是图像的强度。这样一个表达式可以代表一幅活动的、彩色的、立体图像。

当我们研究的是静止图像（Still Image）时，则上式与时间 t 无关；当研究的是单色图像时，显然与波长 λ 无关；对于平面图像来说，则与坐标 z 无关。因此，对于静止的、平面的、单色的图像来说，其数学表达式可以简化为一个二维函数

$$I = f(x, y) \tag{A-2}$$

这里，x 和 y 是二维空间坐标，而函数 f 是求取任意一对二维空间坐标(x, y)上的幅值 I，也就是该点图像的强度或灰度。当 x，y 和幅值 I 为有限的离散数值时，称该图像为数字图像（Digital Image）。数字图像是由有限的元素组成的，每个元素都有一个特定的位置和幅值，这些元素称为图像元素、画面元素或像素。

A.2 数字图像的信息内容

视觉信息是人类获取外部知识、了解世界的主要途径和重要形式。许多情况下，没有任何其他形式比图像所传递的信息更丰富和真切。概括起来，图像信息大致可以分成三类，即符号信息、景物信息和情绪信息[104]。

1. 符号信息

在这类信息中，一般是用文字、符号、图形等表示的具体的或抽象的事物。例如文字，利用文字可组成文章，在某种意义上也可以看成是用二值图像的形式携带这篇文章的寓意。电路图、机械图、建筑图和流程图等，也都是用二值图像的形式向人们提供信息的。因为符号信息是以某一规则进行排列的记号，所以在传送和处理过程中只需表达清楚即可，允许有较大的压缩。

2. 景物信息

这是一种能给人以主观感觉但并不取决于人本身的客观场景信息。一般来讲，它包含丰富的内容，所含的信息量也较多。例如，由生产车间视频监控仪

器上看到的图像信息，可以从中得到有关产品的生产情况、工人的工作情景、设备的运转情况以及车间环境等。情景画面的内容一般比较复杂，需要保留一些细节信息，所以在传输和处理过程中很难进行较大的压缩。

3. 情绪信息

这是一类依赖于观赏者的图像信息，它不仅能给人以直观感觉，而且能以其特殊的艺术内容刺激人的感官，使观赏者“触景生情”引起感情上的波动和情绪上的共鸣。因此，它包含有更多的信息量。例如，我们看到漆黑夜晚、雷电交加的场景时，往往会感到恐惧和敬畏；看到天色阴暗、秋雨绵绵的场景时，一般会有无限的压抑之感；而看到春光明媚、微风和煦的场景时，自然会产生一种轻松欢喜的情绪。这些图像信息不仅取决于图像本身的内容，而且还与观赏者的经历、年龄、嗜好、文化修养以及此时此刻的心境有关，也就是说同一幅图像对观赏者产生的效果是有差异的。

数字图像丰富的信息内容也就决定了图像理论和技术涉及众多的学科，如各类数学、物理学、信号处理、控制论、模式识别、人工智能、生物学、神经心理学、计算机科学与技术等，它是一门兼具交叉性和开放性的学科。

A.3　图像处理的技术门类

目前，数字图像处理多采用计算机处理，因此，有时也称为计算机图像处理（Computer Image Processing）。数字图像处理涉及多个知识门类，具体的方法技术也是种类繁多，应用非常广泛，但从主要研究内容上可以分为以下几个方面：

1. 图像数字化（Image Digitization）

将连续色调的模拟图像经采样量化后转换成数字影像的过程。其目的是将模拟形式的图像通过数字化设备变为数字计算机可用的离散的图像数据，主要包括取样技术和量化技术。

2. 图像变换（Image Transformation）

按一定规则从一帧图像转化生成另一帧图像的处理方法。主要是为了便于后续的工作，采用相关技术以改变图像的表示域和表示数据，主要包括傅里叶变换、余弦变换、沃尔什-哈达玛变换、奇异值分解、KL 变换等。

3. 图像增强（Image Enhancement）

图像增强将原来不清晰的图像变得清晰或强调某些关注的特征，抑制非关注的特征，使之改善图像质量、丰富信息量，加强图像判读和识别效果的图像

处理方法。图像增强技术可分成两大类——频率域法和空间域法。前者把图像看成一种二维信号，对其进行基于二维傅里叶变换的信号增强。采用低通滤波（即只让低频信号通过）法，可去掉图中的噪声；采用高通滤波法，则可增强边缘等高频信号，使模糊的图片变得清晰。具有代表性的空间域算法有局部求平均值法和中值滤波（取局部邻域中的中间像素值）法等，它们可用于去除或减弱噪声。

4. 图像恢复（Image Restoration）

图像恢复也叫图像复原，是通过计算机对质量下降的图像加以重建或恢复的处理过程。因摄像机与物体相对运动、系统误差、畸变、噪声等因素的影响，图像往往不是真实景物的完善映像。在图像恢复中，需建立造成图像质量下降的退化模型，然后运用相反过程来恢复原来图像，并运用一定准则来判定是否得到图像的最佳恢复。尤其是在处理遥感图像时，为消除遥感图像的失真、畸变，恢复目标的反射波谱特性和正确的几何位置，通常需要对图像进行恢复处理，包括辐射校正、大气校正、条带噪声消除、几何校正等内容。

5. 图像分割（Image Segmentation）

图像分割是指根据选定的特征将图像划分为若干个有意义的部分，从而使原图像在内容表达上更为简单明了，为后续图像分析和理解打下基础。传统的图像分割算法按照用户参与的程度可分为自动、交互式与纯手工的分割方法；根据利用区域内相似性还是区域间相异性原理的区别可分为基于区域、基于边界或者两者结合的算法；依据分割结果的确定性与否可以分为软分割与硬分割等。

6. 图像理解（Image Understanding）

图像理解就是对图像的语义理解，有时也叫景物理解。它是以图像为对象，知识为核心，研究图像中有什么目标、目标之间的相互关系、图像是什么场景以及如何应用场景的一门学科。其重点是在图像分析的基础上进一步研究图像中各目标的性质及其相互关系，并得出对图像内容含义的理解以及对原来客观场景的解释，进而指导和规划行为。图像理解所操作的对象是从描述中抽象出来的符号，其处理过程和方法与人类的思维推理有许多相似之处。

7. 图像压缩（Image Data Compression）

图像压缩是指以较少的比特有损或无损地表示原来的像素矩阵的技术，也称图像编码。图像数据之所以能被压缩，就是因为数据中存在着冗余，主要表现为图像中相邻像素间的相关性引起的空间冗余；图像序列中不同帧之间存在

相关性引起的时间冗余；不同彩色平面或频谱带的相关性引起的频谱冗余。数据压缩的目的就是通过去除这些数据冗余来减少表示数据所需的比特数。由于图像数据量的庞大，在存储、传输、处理时非常困难，因此图像数据的压缩就显得非常重要。

8. 图像重建（Image Reconstruction）

图像变换、图像增强、图像恢复都是从图像到图像的处理，即输入的原始数据是图像，处理后输出的也是图像。而图像重建是从数据到图像的处理，也就是说输入的是某种数据，处理结果得到的是图像。图像重建的主要算法有代数法、迭代法、傅里叶反投影法、卷积反投影法等，图像重建的典型应用就是 CT 技术。值得注意的是，三维重建技术与计算机图形学相结合，把多个二维图像合成三维图像，并加以光照模型和各种渲染技术，能生成各种具有强烈真实感及纯净的高质量图像，是虚拟现实和科学可视化技术的基础。

附录 B　模式组合的一些基本概念

B.1　图

图的本质内容是二元关系，图又分为无向图和有向图两种。

定义 B-1（无向图）　无向图 G 定义为一个二元组 $G=(N, E)$，其中，N 是顶点的非空有限集合，$N=\{n_i \mid i=0, \cdots, k\}$；$E$ 是边的有限集合，$E=\{(n_i, n_j) \mid n_i, n_j \in N\}$。

定义 B-2（有向图）　有向图 D 定义为一个二元组 $D=(N, E)$，其中，N 是顶点的非空有限集合，$N=\{n_i \mid i=0, \cdots, k\}$；$E$ 是边的有限集合，$E=\{(n_i, n_j) \mid n_i, n_j \in N\}$ 且 $(n_i, n_j) \neq (n_j, n_i)$，$(n_i, n_j) \in E$ 是顶点 n_i 的出边，顶点 n_j 的入边。

定义 B-3（连通图）　连通图是一个无向图 $G=(N, E)$ 或有向图 $D=(N, E)$，对于 N 中的任意两个顶点 n_s 和 n_t，存在一个顶点的序列 P，使得 $n_s=n_{i_0}, n_{i_1}, \cdots, n_{i_k}=n_t$ 均属于 N，且 $e_j=(n_{i_j}, n_{i_{j+1}})(j=0, 1, \cdots, k-1)$ 均属于 E。P 也被称为图 G 或 D 的一条路径或通路。

定义 B-4（回路）　设 P 是有向图 D 的一条路径，$P=n_{i_0}, n_{i_1}, \cdots, n_{i_k}$，如果 $n_{i_0}=n_{i_k}$，则称 P 是 D 的一条回路，即开始和终结于同一顶点的通路。如果 $k=0$，则 P 称为自回路。若 P 是无向图 G 的一条路径，$P=n_{i_0}, n_{i_1}, \cdots, n_{i_k}$，$n_{i_0}=n_{i_k}$，且 $k>0$，那么，称 P 是 G 的一条回路。若图中无任何回路，则称该图为无回路图。

B.2　树

定义 B-5（树）　一个无回路的无向图称为森林。一个无回路的连通无向图称为树（或自由树）。如果树中有一个节点被特别地标记为根节点，那么这棵树称为根树。

从逻辑结构上讲，树是包含 n 个节点的有穷集合 S（$n>0$），且在 S 上定义了一格关系 R，R 满足以下三个条件：

1）有且仅有一个节点 $t_0 \in S$，该节点对于 R 来说没有前驱，节点 t_0 称作树根；

2）除了节点 t_0 以外，S 中的每个节点对于 R 来说，都有且仅有一个直接前驱；

3）除了节点 t_0 以外的任何节点 $t \in S$，都存在一个节点序列 t_0，t_1，…，t_k，使得 t_0 为树的根，$t_k = t$，有序对 $\langle t_{i-1}, t_i \rangle \in R(1 \leqslant i \leqslant k)$，则该节点序列称为从根节点 t_0 到节点 t 的一条路径。

在根树中，自上而下的路径末端节点称为树的叶节点，介于根节点和叶节点之间的节点称为中间节点（或称内节点）。

在图 B-1 所示的例子中，A 为根节点，C、D、E 为叶节点，B 为中间节点，A 为 B、C 节点的父节点，B、C 称为 A 节点的子节点或后裔，D、E 互为兄弟节点，它们都是 B 节点的子节点。

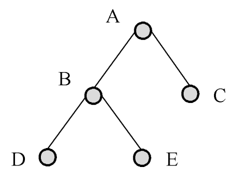


图 B-1　树

B.3　符号串

定义 B-6（符号集）　符号集 Σ 是符号元素的非空有穷集合。典型的符号有字母、数字、各种标点符号和各种运算符。

例如，集合 $\{a, b, c, +, ^*\}$ 是一个含有 5 个符号的符号集，而符号集 $\{0, 1\}$ 只有两个符号。

定义 B-7（符号串）　由符号集 Σ 中 0 个或多个符号相连而成的有穷序列称为 Σ 上的符号串。特殊地，不包括任何符号的符号串称为空串，记作 ε。包括空串在内的 Σ 上符号串的全体记为 Σ^*。

例如，有符号集 $\{a, b, c, +, ^*\}$，则 a，b，c，$+$，*，aa，ab，$a+$，a^*，aaa，$c+^*$ 等等都是该符号集上的符号串。

定义 B-8（符号串的长度）　若 x 是符号集 Σ 上的符号串，那么，其长度指 x 中所含符号的个数，记为 $|x|$。

例如：$|abc| = 3$，$|abc+^*abc| = 8$，而 $|\varepsilon| = 0$。

“连接”和“闭包”是符号串操作中的两种基本运算。

定义 B-9（符号串的连接）　假定 x，y 是符号集 Σ 上的符号串，则把 y 的各个符号依次写在 x 符号串之后得到的符号串称为 x 与 y 的连接，记作 xy。

例如：$\Sigma = \{a, b, c\}$，$x = abc$，$y = cba$，那么，$xy = abccba$。

如果 x 是符号串，把 x 自身连接 $n(n \geqslant 0)$ 次得到的符号串 $z = \overbrace{xx\cdots x}^{n}$，称为 x

的 n 次方幂，记作 x^n。当 $n=0$ 时，$x^0=\varepsilon$。当 $n\geqslant 1$ 时，$x^n=xx^{n-1}=x^{n-1}x$。

定义 B-10（集合的乘积运算）　设 A，B 是符号集 Σ 上的两个符号串集合，则 A 和 B 的乘积定义为

$$AB=\{xy \mid x\in A,\ y\in B\} \tag{B-1}$$

其中，$A^0=\{\varepsilon\}$。当 $n\geqslant 1$ 时，$A^n=A^{n-1}A=AA^{n-1}$。

定义 B-11（集合的闭包运算）　设 V 是符号集 Σ 上的一个符号串集合，则 V 的正闭包定义为

$$V^+=V^1\cup V^2\cup\cdots\cup V^n\cup\cdots \tag{B-2}$$

V 的闭包定义为

$$V^*=V^0\cup V^+ \tag{B-3}$$

例如：$V=\{a,\ b\}$，则

$$V^+=\{a,\ b,\ aa,\ ab,\ ba,\ bb,\ aaa,\ aab,\ \cdots\}$$
$$V^*=\{\varepsilon,\ a,\ b,\ aa,\ ab,\ ba,\ bb,\ aaa,\ aab,\ \cdots\}$$

附录 C　概率统计的一些预备知识

C.1　概率

概率（Probability）是从随机实验 E 中的事件到实数域的映射函数，用以表示事件发生的可能性。如果用 $P(A)$作为事件 A 的概率，S 是实验的样本空间，则概率函数必须满足如下三条公理：

公理 C-1（非负性）　$0 \leqslant P(A) \leqslant 1$

公理 C-2（规范性）　$P(S)=1$

公理 C-3（可列可加性）　如果事件 A_1，A_2，…，A_m，…两两互斥，即对于任意的 i 和 j（$i \neq j$），事件 A_i和 A_j不相交（$A_i \cap A_j = \varnothing$），则有

$$P(A_1 \cup A_2 \cup \cdots \cup A_m \cup \cdots) = P(A_1) + P(A_2) + \cdots + P(A_m) + \cdots \qquad \text{(C-1)}$$

C.2　最大似然估计

如果 $S=\{s_1, s_2, \cdots, s_n\}$是一个随机实验 E 的样本空间，在相同的情况下重复实验 N 次，观察到样本 $s_k(1 \leqslant k \leqslant n)$的次数为 $n_N(s_k)$，那么，s_k在这 N 次实验中的相对频率为

$$q_N(s_k) = \frac{n_N(s_k)}{N} \qquad \text{(C-2)}$$

由于 $\sum_{k=1}^{n} n_N(s_k) = N$，因此，$\sum_{k=1}^{n} q_N(s_k) = 1$。

当 N 越来越大时，相对频率 $q_N(s_k)$就越来越接近 s_k的概率 $P(s_k)$。事实上，

$$\lim_{N \to \infty} q_N(s_k) = P(s_k) \qquad \text{(C-3)}$$

因此，通常用相对频率作为概率的估计值，这种估计概率值的方法称为最大似然估计（Likelihood Estimation）。

C.3　条件概率

如果 A 和 B 是样本空间 S 上的两个事件，$P(B)>0$，那么，在给定 B 时 A

的条件概率（Conditional Probability）$P(A|B)$为

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{C-4}$$

条件概率 $P(A|B)$给出了在已知事件 B 发生的情况下，事件 A 的概率。一般地，$P(A|B)\neq P(A)$，$P(AB)$即为 $P(A\cap B)$。

根据公式（C-4），有

$$P(AB)=P(B)P(A|B)=P(A)P(B|A) \tag{C-5}$$

这个等式有时称为概率的乘法定理或乘法规则，其一般形式表示为

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{C-6}$$

条件概率也有三个基本性质：

（1）非负性：$P(A|B)\geqslant 0$

（2）规范性：$P(S|B)=1$

（3）可列可加性：如果事件 A_1，A_2，…，A_m，…两两互斥，则有

$$P(A_1\cup A_2\cup\cdots\cup A_m\cup\cdots|B)=P(A_1|B)+P(A_2|B)+\cdots+P(A_m|B)+\cdots \tag{C-7}$$

C.4 贝叶斯公式

贝叶斯公式，或称逆概率公式，是条件概率计算的重要依据。实际上，根据条件概率的定义公式（C-4）和乘法规则式（C-5），可得

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B|A)}{P(B)} \tag{C-8}$$

式（C-8）右边的分母可以看成一个常量，因为我们只关心在给定事件 B 的情况下可能发生事件 A 的概率，$P(B)$的值是确定不变的，下面给出它的计算方法。

定理 C-1（全概率公式） 如果事件 A_1，A_2，…，A_n满足：

（1）A_1，A_2，…，A_n两两互斥，且 $P(A_i)>0$，$(i=1, 2, \cdots, n)$；

（2）$A_1\cup A_2\cup\cdots\cup A_n=S$（完全性）

则对任何事件 B 有

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i) \tag{C-9}$$

由定理 C-1，我们可以修改公式（C-8），进而给出贝叶斯公式。

定理 C-2（贝叶斯公式） 设事件 A_1，A_2，…，A_n满足定理 C-1 的条件（1）（2）。则对任何事件 B，当 $P(B)>0$ 有

$$P(A_j \mid B) = \frac{P(A_j)P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)} \tag{C-10}$$

其中，i，$j=1$，2，3，…，n。

C.5　随机变量

一个随机实验可能有多种不同的结果，到底会出现哪一种，存在一定的概率。简单地说，随机变量（Random Variable）就是实验结果的函数。设离散型随机变量 X 的所有可能值为 x_k，$k=1$，2，3，…，n，…，X 取各可能值的概率为

$$P[X=x_k]=p_k,\ k=1,2,3,\cdots,n,\cdots \tag{C-11}$$

且 p_k满足 $p_k \geqslant 0$（非负性）与 $\sum_{k=1}^{\infty} p_k = 1$（归一性），则称式（C-11）为离散型随机变量 X 的概率分布或分布律。(也称分布列，分布密度)。此时，函数

$$F(x)=P[X \leqslant x],\ -\infty < x < \infty \tag{C-12}$$

称为 X 的分布函数。

C.6　二项式分布

有一类广泛存在的实验，其特点是只有对立的两个结果，即实验 E 的样本空间只有两个基本事件 A 与 $\overline{A}$，我们称之为伯努利实验。将伯努利实验独立重复进行 n 次，则称 n 重伯努利实验，这里所谓“重复”是指每次实验条件相同，事件 A 发生的概率 $P(A)=p$ 保持不变。

一般，如果离散型随机变量 X 的分布律为

$$p[X=k]=C_n^k p^k q^{n-k},\ k=1,2,3,\cdots,n(0<p<1) \tag{C-13}$$

则称 X 服从参数是 n，p 的二项式分布（Binomial Distribution），并记成 $X \sim B(n,\ p)$。在 n 重伯努利实验中，若 $P(A)=p$，则 A 发生的次数 X 服从参数是 n，p 的二项式分布。

二项式分布是最重要的离散型概率分布之一。例如，在图像处理中如果以局部特征为处理单位，为了简化问题的复杂性，通常假设一个局部特征的出现独立于其他局部特征，这样一来，局部特征的概率分布就近似地被认为符合二项式分布。

C.7　联合概率分布和条件概率分布

若二维随机变量（X，Y）所有可能取值$(x,\ y)$只有有限个或可列多个，则

称（X，Y）为二维离散型随机变量，其联合概率分布（Joint Distribution）为

$$p_{ij}=P[X=x_i, Y=y_j],\ i, j=1, 2, 3, \cdots \tag{C-14}$$

考虑分量 X 在给定 $Y=y_j$ 条件下的概率分布，实际上就是求条件概率

$$P[X=x_i \mid Y=y_j]=\frac{P[X=x_i, Y=y_j]}{P[Y=y_j]}=\frac{p_{ij}}{P[Y=y_j]}=\frac{p_{ij}}{\sum\limits_{i=1}^{\infty}p_{ij}} \tag{C-15}$$

其中，$P[Y=y_j]=\sum\limits_{i=1}^{\infty}p_{ij}$ 是 (X, Y) 关于 Y 的边缘分布律。

类似的，在 $X=x_i$ 条件下，分量 Y 的条件分布律为

$$P[Y=y_j \mid X=x_i]=\frac{p_{ij}}{\sum\limits_{j=1}^{\infty}p_{ij}} \tag{C-16}$$

其中，$P[X=x_i]=\sum\limits_{j=1}^{\infty}p_{ij}$ 是 (X,Y) 关于 X 的边缘分布律。

C.8　贝叶斯决策理论

贝叶斯决策理论（Bayesian Decision Theory）是统计方法处理模式分类问题的基本理论之一。假设研究的分类问题有 N 个类别，每个类别 ω_i（$i=1$，2，…，N）出现的先验概率为 $P(\omega_i)$。在特征空间已经观察到某个特定的模式 $\mathbf{x}$，且条件概率密度函数 $p(\mathbf{x} \mid \omega_i)$ 是已知的。那么，利用贝叶斯公式可以得到后验概率

$$P(\omega_i \mid \mathbf{x})=\frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{\sum\limits_{j=1}^{n}p(\mathbf{x} \mid \omega_j)P(\omega_j)} \tag{C-17}$$

基于最小错误率的贝叶斯决策规则为：如果 $P(\omega_i \mid \mathbf{x})=\max\limits_{j=1,2,\cdots,N}P(\omega_j \mid \mathbf{x})$，也就是说，如果 $p(\mathbf{x} \mid \omega_i)P(\omega_i)=\max\limits_{j=1,2,\cdots,N}p(\mathbf{x} \mid \omega_j)P(\omega_j)$，那么将模式 $\mathbf{x}$ 赋予类 ω_i，即 $\mathbf{x}\in\omega_i$。

上述理论中，每个类的出现概率以模式的条件概率密度函数必须是已知的。前者的获取通常并不构成问题，比如，当所有类的出现概率大致相同，则可令 $P(\omega_i)=1/N$，即使这个条件不正确，我们也可以通过对问题的认识推算出这些先验概率。而后者的估计就是另一回事了，如果模式向量 $\mathbf{x}$ 是 n 维的，那么 $p(\mathbf{x} \mid \omega_i)$ 就是一个 n 元函数，如果它的形式是未知的，就需要使用多元概率理论的方法对它进行估计。这类方法在实际应用中非常困难，尤其是代表每个类别的模式数目不大，或隐含的概率密度函数形式的规律性不强时更是如此。由于这些原因，贝叶斯决策理论在实际应用中通常要假设各种概率密度函数的解

析式，以及从每类样本模式估计的必要参数。目前，对 $p(\mathbf{x} \mid \omega_i)$的最为普遍的假设形式是高斯概率密度函数[8]。

C.9　期望和方差

期望值（Expectation）是指随机变量所取的概率平均。假设 X 为一个随机变量，其概率分布为 $P[X = x_k] = p_k$，$k = 1$，2，3，…，n，…，若级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛,则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 为随机变量 X 的数学期望或均值,记作 $E(X)$,即

$$E(X) = \sum_{k=1}^{\infty} x_k p_k \tag{C-18}$$

一个随机变量的方差（Variance）描述的是该随机变量的值偏离其期望值的程度。设 X 为一个随机变量，那么它的方差为

$$D(X) = E[X - E(X)]^2 = E(X^2) - E^2(X) \tag{C-19}$$

称 $D(X)$的正方根 $\sqrt{D(X)}$为随机变量 X 的标准差或均方差，记为

$$\sigma(X) = \sqrt{D(X)} \tag{C-20}$$

$\sigma(X)$也描述随机变量 X 取值的离散程度，可简记为 σ，$D(X)$也可简记为 σ^2。

附录 D　信息检索的一些基础模型

信息检索（Information Retrieval，IR）的研究起源于图书馆的资料查询和文摘索引工作。计算机诞生以后，尤其是随着计算机网络技术的迅速发展，信息检索的内容已经从传统的文本检索扩展到包含图片、音频、视频等多媒体信息的检索；检索对象从相对封闭、稳定一致、由独立数据库集中管理的信息内容扩展到开放、动态、更新速度快、分布广泛、管理松散的网络内容；信息检索的用户由原来的情报专业人员扩展到包括商务人员、管理人员、教师、学生、各专业技术人员等在内的普通大众[129]。

海量互联网信息的涌现是信息检索技术发展最直接的驱动力，这对信息检索从结果到方式都提出了更高、更多样化的要求。而信息检索研究的目的是寻找从资料中获取可用信息的模型和算法，所以无论检索内容如何丰富、如何变化，其本质还是一样的。我们下面还是以传统的文档资料检索为例，介绍一些基础的、成熟的模型（不妨称之为“检索模型”），这些模型已经在多媒体信息检索中广为借鉴。

D.1　布尔模型

在这种模型中，候选查询文档 D 由关键词的逻辑组合表达式表示，用户查询 Q 由布尔表达式表示，那么，相关度 $R(D, Q) = D \rightarrow Q$，即当 $D \rightarrow Q$ 成立时，$R(D, Q) = 1$，否则，$R(D, Q) = 0$。

例如：$D = \text{computer} \wedge \text{graphics} \wedge \text{interface} \wedge \text{user}$，$Q = \text{computer} \wedge (\text{graphics} \vee \text{interface})$，if$D \rightarrow Q$then$R(D, Q) = 1$。

这种方法的主要问题是，相关度为二值逻辑，要么为 1，要么为 0。也就是说，候选文档与用户查询语句要么相关，要么不相关，这在实际情况下是不合理的。另外，作为一般的终端用户，很难快速正确地给出查询语句的布尔表达式。

D.2　向量空间模型

向量空间模型的基本思想是：整个向量空间由关键词构成，即

$\langle t_1, t_2, \cdots, t_n\rangle$；候选文档 $D=\langle a_1, a_2, \cdots, a_n\rangle$，其中，$a_i$（$1\leqslant i\leqslant n$）为 D 中 t_i 的权重；用户查询语句 $Q=\langle b_1, b_2, \cdots, b_n\rangle$，其中，$b_i$（$1\leqslant i\leqslant n$）为 Q 中 t_i 的权重。那么用户查询与候选文档的相关度 $R(D, Q)=Sim(D, Q)$ 可以由以下方法求得：

（1）点积法

$$Sim(D, Q)=D\cdot Q=\sum_i (a_i\times b_i) \tag{D-1}$$

（2）余弦法

$$Sim(D, Q)=\frac{D\cdot Q}{\|D\|\times\|Q\|}=\frac{\sum_i (a_i\times b_i)}{\sqrt{\left(\sum_i a_i^2\right)\left(\sum_i b_i^2\right)}} \tag{D-2}$$

（3）Dice 方法

$$Sim(D, Q)=\frac{2\times D\cdot Q}{\|D\|^2+\|Q\|^2}=\frac{2\sum_i (a_i\times b_i)}{\sum_i a_i^2+\sum_i b_i^2} \tag{D-3}$$

（4）Jaccard 方法

$$Sim(D, Q)=\frac{D\cdot Q}{\|D\|^2+\|Q\|^2-D\cdot Q}=\frac{\sum_i (a_i\times b_i)}{\sum_i a_i^2+\sum_i b_i^2-\sum_i (a_i\times b_i)} \tag{D-4}$$

度量两个向量之间的相似性，还有很多方法，在此就不一一列举，可以参阅相关资料，如参考文献［54，68，129］等等。

D.3　概率模型

概率模型的基本思想是：给定查询语句 Q，候选文档 D，用 R 表示 D 与 Q 相关，$\overline{R}$ 表示 D 与 Q 不相关，那么，可以根据概率 $P(R\mid D, Q)$ 和 $P(\overline{R}\mid D, Q)$ 这两个值的大小选取搜索的文档。

根据贝叶斯公式：

$$P(R|D, Q)=\frac{P(D|R, Q)\times P(R, Q)}{P(D, Q)}\propto P(D|R, Q)=P(D|R_Q) \tag{D-5}$$

假定文档 $D=\langle x_1, x_2, \cdots, x_n\rangle$，其中，$x_i=\begin{cases}1，关键词\ t_i 出现\\ 0，关键词\ t_i 不出现\end{cases}$，那么

$$\begin{aligned}P(D\mid R, Q)&=\prod_{x_i\in D}P(x_i\mid R_Q)\\&=\prod_{t_i}P(x_i=1\mid R_Q)^{x_i}P(x_i=0\mid R_Q)^{(1-x_i)}\\&=\prod_{t_i}p_i^{x_i}(1-p_i)^{(1-x_i)}\end{aligned} \tag{D-6}$$

$$P(D \mid \bar{R}, Q) = \prod_{t_i} P(x_i = 1 \mid \bar{R}_Q)^{x_i} P(x_i = 0 \mid \bar{R}_Q)^{(1-x_i)}$$
$$= \prod_{t_i} q_i^{x_i} (1 - q_i)^{(1-x_i)} \tag{D-7}$$

文档与查询的相关度：

$$R(D, Q) = \log \frac{P(D \mid R, Q)}{P(D \mid \bar{R}, Q)} = \log \frac{\prod_{t_i} p_i^{x_i}(1-p_i)^{(1-x_i)}}{\prod_{t_i} q_i^{x_i}(1-q_i)^{(1-x_i)}}$$
$$= \sum_{t_i} x_i \log \frac{p_i(1-q_i)}{q_i(1-p_i)} + \sum_{t_i} \log \frac{1-p_i}{1-q_i}$$
$$\propto \sum_{t_i} x_i \log \frac{p_i(1-q_i)}{q_i(1-p_i)} \tag{D-8}$$

余下的问题就是如何估计概率 $p_i = P(x_i = 1 \mid R_Q)$和 $q_i = P(x_i = 1 \mid \bar{R}_Q)$。

假设一组训练样本共有 N 个文档，其中，R_i个与查询 Q 相关的文档，$N - R_i$个不相关的文档，这 N 个文档中有 n_i个文档包含关键词 t_i。R_i个相关文档中有 r_i个文档包含关键词 t_i，$R_i - r_i$个文档不包含关键词 t_i；$N - R_i$个不相关的文档中有 $n_i - r_i$个文档包含关键词 t_i，$N - R_i - n_i + r_i$个不包含关键词 t_i。如表 D-1 所示。

表 D-1　训练样本数目关系

	相关文档	不相关文档
数量	R_i	$N - R_i$
包含 t_i的文档数	r_i	$n_i - r_i$
不包含 t_i的文档数	$R_i - r_i$	$N - R_i - n_i + r_i$

那么

$$p_i = \frac{r_i}{R_i} \tag{D-9}$$

$$q_i = \frac{n_i - r_i}{N - R_i} \tag{D-10}$$

于是，公式（D-8）可以进一步改写为

$$R(D, Q) = \sum_{t_i} x_i \log \frac{p_i(1-q_i)}{q_i(1-p_i)} = \sum_{t_i} x_i \log \frac{r_i(N - R_i - n_i + r_i)}{(R_i - r_i)(n_i - r_i)} \tag{D-11}$$

概率模型在理论上具有较好的数学基础，但是，在不进行任何简化的情况下，实现起来比较困难，其有效性往往受到诸多因素的影响。

D.4　语言模型

鉴于语言模型在很多问题的研究中都获得了成功的应用，很多学者也提出

了将改进的语言模型用于信息检索的方法。例如，文档模型（Document Model）、查询模型（Query Model）、差异模型（Divergence Model）和翻译模型（Translation Model）等。

文档模型的基本思想是：假定查询 Q 是由文档 D 的概率模型产生的，并由此对文档进行排序。也就是说，给定查询 $Q=q_1q_2\cdots q_n$（q_i为查询词）和文档 D，那么，文档模型的任务就是先建立文档的语言模型 M_D，然后根据概率 $P(Q\mid M_D)$对文档进行排序。

文档模型的一元文法描述形式为

$$P(Q|M_D)=\prod_{q_i\in Q}P(q_i|M_D) \tag{D-12}$$

$P(q_i\mid M_D)$反映的是查询词在文档 D 中的概率分布。

查询模型的基本思想是：假定查询 $Q=q_1q_2\cdots q_n$和文档 D 均采样自一个未知的相关模型 R，R 刻画了 Q 和 D 在查询相关文档中的概率分布；从相关模型 R 中经过 k 次采样，观察到查询 Q，估计第 $k+1$ 次采样观察到文档中的词 ω 的概率。

查询模型描述为

$$P(D|R)=\prod_{\omega\in D}P(\omega|R) \tag{D-13}$$

$$P(\omega|R)\approx P(\omega|q_1q_2\cdots q_n)=\frac{P(\omega,q_1q_2\cdots q_n)}{P(q_1q_2\cdots q_n)} \tag{D-14}$$

差异模型的基本思想是：通过计算文档模型和查询模型之间的 Kullback-Leibler 差异（KL 距离），根据 KL 距离大小对候选文档进行排序。那么，该模型的任务就是先估计文档模型 $P(\omega\mid M_D)$，然后估计查询模型 $P(\omega\mid R)$，从而计算文档模型和查询模型之间的 KL 距离：

$$\mathrm{KL}(R\,||\,M_D)=\sum_{\omega}P(\omega\mid R)\log\frac{P(\omega\mid R)}{P(\omega\mid M_D)} \tag{D-15}$$

翻译模型的基本思想是：把查询语句 $Q=q_1q_2\cdots q_n$看做是文档 D 在同一语言内的翻译，并根据翻译的概率大小对候选文档进行排序，根据统计翻译模型有

$$P(Q\mid D)=\prod_{i}P(q_i\mid D)=\prod_{i}\sum_{j}P(q_i\mid\omega_j)P(\omega_j\mid D) \tag{D-16}$$

其中，$P(\omega_j\mid D)$为词 ω_j在文档 D 中的概率分布，$P(q_i\mid\omega_j)$为词 ω_j翻译成查询中的词 q_i的概率。

附录 E　名词术语解释

本附录旨在避免读者对常用词和本书所使用的专业化词汇产生混淆，方便读者对本书的阅读和理解。下述解释同后面参考文献中所列的已经出版的图像处理和计算机技术方面的书籍中对有关词汇的定义大体一致，但不一定都是本领域的标准化定义，敬请注意。

10-fold cross-validation，十折交叉验证——常用的精度测试方法，将数据集分成 10 份，轮流将其中 9 份做训练，1 份做测试，10 次结果的均值作为对算法精度的估计。

Active Contour Model，主动轮廓模型——又被称为 Snake，是由 Andrew Blake 教授提出的一种目标轮廓描述方法，主要应用于基于形状的目标分割。

Artificial Neural Networks，人工神经网络——简称神经网络（NN/NNet/ANN）或称作连接模型（Connection Model），是一种模仿动物神经网络行为特征，进行分布式并行信息处理的模型。

Binary image，二值图像——只有两级灰度的数字图像（通常为 0 和 1，黑和白）。

Boundary chain code，边界链码——定义一个物体边界的方向序列。

Boundary pixel，边界像素——至少和一个背景像素相邻接的内部像素。

Boundary tracking，边界跟踪——一种图像分割技术，通过沿弧从一个像素顺序探索到下一个像素的方法将弧检测出来。

Brightness，亮度——和图像一个点相关的值，表示从改点的物体发射或反射的光的量。

Cluster，聚类，集群——在空间（如特征空间）中位置接近的点的集合。

Cluster analysis，聚类分析——在空间中对聚类的检测、度量和描述。

Computer-assisted diagnosis，计算机辅助诊断——英文简称 CAD ，是指通过影像学、医学图像处理技术以及其他可能的生理、生化手段，结合计算机的分析计算，辅助影像科医师发现病灶，提高诊断的准确率。

Contrast，对比度——物体平均亮度（或灰度）与其周围背景的差别程度。

Curve，曲线——（1）空间的一条连续路径；（2）表示一路径的像素集合。

Degree of freedom，自由度——能够自由取值的变量个数，如有 3 个变量 x、y、z，但限制条件为 $x + y + z = 18$，因此其自由度为 2。

Digital image，数字图像——见附录 A. 1。

Digital image processing，数字图像处理——对图像的数字化处理；由计算机对图像信息进行操作。

Digitization，数字化——将景物图像转化为数字形式的过程。

Edge，边缘——（1）在图像中灰度出现突变的区域；（2）属于一段弧上的像素集，在其另一边的像素与其有明显的灰度差别。

Edge detection，边缘检测——通过检查邻域，将边缘像素标识出的一种图像分割技术。

Edge enhancement，边缘增强——通过将边缘两边像素的对比度扩大来锐化图像边缘的一种图像处理技术。

Enhance，增强——增加对比度或主观可视程度。

Face recognition，人脸识别——指利用分析比较人脸视觉特征信息进行身份鉴别的计算机技术。

False negative，负误识——在二分类模式识别中，将属于目标标注为不属于目标的误分类。

False positive，正误识——在二分类模式识别中，将不属于目标标注为属于目标的误分类。

Feature，特征——物体的一种特性，它可以度量。

Feature extraction，特征检测——模式识别过程中的一个步骤，在该步骤中计算物体的有关度量。

Featureselection，特征选择——对原始特征进行筛选，舍弃那些对类别区分并无多大贡献的特征，使得最终的特征空间能够反映分类的本质。

Feature space，特征空间——即度量空间，在模式识别中，包含所有可能度量向量的 n 维向量空间。

Fourier transform，傅里叶变换——采用复指数 $e^{-j2\pi sx} = \cos(2\pi sx) + j\sin(2\pi sx)$ 作为核函数的一种线性变换。

Geometric correction，几何校正——采用几何变换消除几何畸变的一种图像复原技术。

Gray level，灰度级——（1）和数字图像的像素相关联的值，它表示由该像素的原始景物点的亮度；（2）在某像素位置对图像的局部性质的数字化度量。

Gray scale，灰度——在数字图像中所有可能灰度级的集合。

Gray-scale transformation，灰度变换——在点运算中的一种函数，它建立了输入灰度和对应输出灰度的关系。

Image，图像——对物理景物或其他图像的统一表示，见附录 A. 1。

Image compression，图像压缩——消除图像冗余或对图像近似的一种过程，其目的是让图像以更紧凑的形式表示。

Image coding，图像编码——将图像变换成另一个可恢复的形式（如压缩）。

Image enhancement，图像增强——旨在提高图像视觉外观的处理方法。

Image matching，图像匹配——为决定两幅图像相似程度对它们进行量化比较的过程。

Image-processing operation，图像处理运算——将输入图像变换为输出图像的一系列步骤。

Image reconstruction，图像重构——从非图像形式构造或恢复图像的过程。

Image registration，图像配准——通过将同一景物的一幅图像和另一幅图像进行几何运算，以使其中物体对准的过程。

Image restoration，图像恢复——通过逆图像退化的方法将图像恢复为原始状态的过程。

Image segmentation，图像分割——（1）在图像中检测并勾画出感兴趣物体的处理；（2）将图像分为不相连的区域，通常这些区域对应于物体以及物体所处的背景。

Information Retrieval，信息检索——指将信息按一定的方式组织起来，并根据信息用户的需要找出有关的信息的过程和技术。

Information theory，信息论——关于信息量度量和信息编码、信号处理和分析的科学理论。

Interior pixel，内像素——在一幅二值图像中，处于物体内部的像素（相对于边界像素、外像素）。

Line detection，线检测——通过检查邻域将直线像素标识出来的一种图像分割技术。

Local property，局部特性——在图像中随位置变化的感兴趣的特性（如光学图像的亮度或颜色，非光学图像的高度、温度和密度等）。

Magnetic resonance imaging，磁共振成像——又称核磁共振成像术，英文简称 MRI。利用人体组织中氢原子核（质子）在磁场中受到射频脉冲的激励而发生核磁共振现象，产生磁共振信号，经过电子计算机处理，重建出人体某一层面的图像的成像技术。

Misclassification，误分类——在模式识别中，将目标错误地标识为其他类别。

Multispectral image，多光谱图像——同一景物的一组图像，每一幅是由电磁谱的不同波段辐射产生的。

Neighborhood，邻域——在给定像素附近的一个像素集合。

Neighborhood operation，邻域运算——基于输入像素的一个邻域的像素灰度决定该像素输出灰度的图像处理运算。

Noise，噪声——一幅图像中阻碍感兴趣数据的识别和解释的不相关部分。

Object，目标，物体——在模式识别中，处于一幅二值图像中的相连像素的集合，通常对应于该图像所表示景物中的一个物体。

Pattern，模式——一个类的成员所表现出的共有的有意义的规则性，可以度量并可用于对感兴趣的目标进行分类。

Pattern class，模式类——可预先赋予一个目标的相互不包容的任一个类别标签。

Pattern classification，模式分类——将目标赋予模式类的过程。

Pattern recognition，模式识别——自动或半自动地检测、度量、分类目标物体。

Perimeter，周长——围绕一个物体的边界的周边距离。

Picture element，图像元素，像素——数字图像的最小基本组成单位。

Pixel，像素——图像元素（picture element）的缩写。

Quantization，量化——在每个像素处，将图像的局部特性赋予一个灰度集合中的元素的过程。

Region，区域——一幅图像中的相连子集。

Region growing，区域生长，区域增长——通过反复对具有相似灰度或纹理的相邻子区域求并集生成区域的一种图像分割技术。

Registered images，已配准图像——同一景物的两幅（或以上）图像已相互调准好位置，从而使其中的物体具有相同的图像位置。

Resolution，分辨率——（1）在光学中指可分辨的点物体之间最小的分离

距离；（2）在图像处理中，指图像中相邻的点物体能够被分辨出的程度。

Scene，场景——客观物体的一种特色布局。

Sharp，清晰——关于图像细节的易分辨性。

Sharpening，锐化——用以增强图像细节的一种图像处理技术。

Smoothing，平滑——降低图像细节幅度的一种图像处理技术，通常用于降噪。

Statistical pattern recognition，统计模式识别——基于概率统计理论，将目标赋予模式类的一种模式识别方法

Structural pattern recognition，结构模式识别——为描述和分类目标，将目标表示为基元及其相互关系的一种模式识别方法。

Syntactic pattern recognition，句法模式识别——采用自然或人工语言模式定义基元及相互关系的一种结构模式识别方法。

Synthetic aperture radar，合成孔径雷达——是一种高分辨率的二维微波对地成像系统，能够全天候工作，有效地识别伪装和穿透掩盖物。

System，系统——对输入作出响应，并生成输出。

Texture，纹理——在图像处理中，表示图像中灰度幅度及其局部变化的空间组织的一种属性。

Threshold，阈值——用以产生二值图像的一个特定的灰度（临界值）。

Thresholding，二值化——由灰度图像产生二值图像的过程，一般如果输入像素的灰度值大于给定的阈值则输出像素赋值为 1，否则赋值为 0。

Virtual Reality，虚拟现实——又称灵境技术或人工环境，英文简称 VR。是利用电脑产生一个三维空间的虚拟世界，提供使用者关于视觉、听觉、触觉等感官的模拟，让使用者如同身临其境一般。

Watershed algorithm，分水岭算法——一种基于拓扑理论的数学形态学的图像分割方法。

参 考 文 献

［1］袁晓辉，金立左，李久贤等．基于兴趣区检测与分析的水上桥梁识别［J］．红外与毫米波学报，2003，22（5）：331-336.

［2］Ong S C W，Ranganath S. Automatic sign language analysis：a survey and the future beyond lexical meaning［J］．IEEE Transactions on Pattern Analysis and Machine Intelligence，2005，27（6）：873-891.

［3］Viola P，Jones M J. Robust real-time face detection［J］．International Journal of Computer Vision，2004，57（2）：137-154.

［4］高峰，文贡坚，吕金建．基于干线对的红外与可见光最优图像配准算法［J］．计算机学报，2007，30（6）：1014-1021.

［5］白静，侯彪，王爽，等．基于提升 Directionlet 域高斯混合尺度模型的 SAR 图像噪声抑制［J］．计算机学报，2008，31（7）：1234-1241.

［6］陈尔学，李增元，田昕，等．尺度不变特征变换法在 SAR 影像匹配中的应用［J］．自动化学报，2008，34（8）：861-868.

［7］孙即祥．图像分析［M］．北京：科学出版社，2005.

［8］冈萨雷斯，伍兹．数字图像处理［M］．阮秋琦，等译．2 版．北京：电子工业出版社，2007.

［9］徐宗本．计算智能——模拟进化计算［M］．北京：高等教育出版社，2004.

［10］李波，郑锦，孟勃．数字媒体内容理解［J］．中国计算机学会通讯，2011，7（2）：16-21.

［11］Li W，Piech V，Gilbert C D. Learning to link visual contours. Neuron［J］．2008，57（3）：442-451.

［12］艾森克 M W，基恩 M T．认知心理学［M］．高定国，等译．上海：华东师范大学出版社，2009.

［13］加洛蒂．认知心理学［M］．吴国宏等译．西安：陕西师范大学出版社，2005.

［14］Marr D. Vision：a computational investigation into the human representation and processing of visual information［M］．Freeman W. H. and Company，San Francisco，1982.

［15］Sharon E，Brandt A，Basri R. Fast multiscale image segmentation［C］．Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition，2000：70-77.

［16］Malik J，Belongie S，Leung，T，et al. Contour and texture analysis for image segmentation［J］．International Journal of Computer Vision，2001，43（1）：7-27.

［17］David P，DeMenthon D. Object Recognition in High Clutter Images Using Line Features［C］．Proceedings of the IEEE International Conference on Computer Vision，2005，2：1581-1588.

[18] Borenstein E, Ullman S. Class-specific, top-down segmentation [C]. Proceedings of the European Conference on Computer Vision, 2002: 639-641.

[19] Yu S X, Shi J. Object-specific figure-ground segregation [C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, 2: 39-45.

[20] Tu Z W, Chen X R, Yuille A L, et al. Image parsing: Unifying segmentation, detection, and recognition [J]. International Journal of Computer Vision, 2005, 63 (2): 113-140.

[21] Ferrari V, Tuytelaars T, Gool L V. Simultaneous recognition and segmentation by image exploration [J]. LNCS 4170: Toward Category-Level Object Recognition. Berlin / Heidelberg: Springer, 2006: 145-169.

[22] Borenstein E, Ullman S. Combining top-down/bottom-up segmentations [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30 (12): 2109-2125.

[23] 卡斯尔曼. 数字图像处理 [M]. 朱志刚, 等译. 北京: 电子工业出版社, 2002.

[24] Russell B C, Torralba A, Murphy K P, et al. A database and webbased tool for image annotation [J]. International Journal of Computer Vision, 2008, 77: 157-173.

[25] Yao B, Yang X, Zhu S C. Introduction to a largescale general purpose ground truth database: Methodology, annotation tool and benchmarks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007: 169-183.

[26] Fergus R, Perona P, Zisserman A. A visual category filter for google images [C]. Proceedings of the European Conference on Computer Vision, 2004: 242-256.

[27] Fergus R, Li F F, Perona P, et al. Learning object categories from google's image search [C]. Proceedings of the IEEE International Conference on Computer Vision, 2005: 1816-1823.

[28] Berg T L, Forsyth D A. Animals on the web [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006: 1463-1470.

[29] Schro F, Criminisi A, Zisserman A. Harvesting image databases from the web [C]. Proceedings of the IEEE International Conference on Computer Vision, 2007: 1-8.

[30] Collins B, Deng J, Li K, et al. Towards scalable dataset construction: An active learning approach [C]. Proceedings of the European Conference on Computer Vision, 2008: 86-98.

[31] Theodoridis S, Koutroumbas K. Pattern Recognition: Second Edition [M]. New York: Academic Press, 2003.

[32] Moosmann F, Nowak E, Jurie F. Randomized clustering forests for image classification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30 (9): 1632-1646.

[33] Heitz G, Elidan G, Packer B, Koller D. Shape-based object localization for descriptive classification [J]. International Journal of Computer Vision, 2009, 84 (1): 40-62.

[34] Leordeanu M, Heber M, Sukthankar R. Beyond local appearance: category recognition from pairwise interactions of simple features [C]. Proceedings of the IEEE Conference on Computer

Vision and Pattern Recognition, 2007: 1-8.

[35] Stark M, Schiele B. How good are local features for classes of geometric objects [C]. Proceedings of the IEEE International Conference on Computer Vision, 2007: 1-8.

[36] Wang Y Z, Zhu S C. Perceptual Scale-space and its applications [J]. International Journal of Computer Vision, 2008, 80 (1): 143-165.

[37] Wu Y N, Guo C E, Zhu S C. From information scaling of natural images to regimes of statistical models [J]. Quarterly of Applied Mathematics, 2008, 66 (1): 81-122.

[38] Shen H F, Zhang L P, Huang B, et al. A MAP approach for joint motion estimation, segmentation, and super resolution [J]. IEEE Transactions on Image Processing, 2007, 16 (2): 479-490.

[39] Flitti F, Collet C. Markovian regularization of latent-variable-models mixture for new multi-component image reduction/segmentation scheme [J]. Signal, Image and Video Processing, 2007, 1 (3): 191-201.

[40] 谷春亮, 尹宝才, 孔德慧, 等. 基于三维多分辨率模型与 Fisher 线性判别的人脸识别方法 [J]. 计算机学报, 2005, 28 (1): 97-104.

[41] 焦李成, 孙强. 多尺度变换域图像的感知与识别: 进展和展望 [J]. 计算机学报, 2006, 29 (2): 177-193.

[42] 席学强. 基于模型的遥感图像三维目标识别系统研究 [D]. 长沙: 国防科技大学, 2000.

[43] 陈晓飞, 王润生. 目标骨架的多尺度树表示 [J]. 计算机学报, 2004, 27 (11): 1540-1545.

[44] Polikar R, Udpa L, Udpa A S, et al. Learn++: an incremental learning algorithm for supervised neural networks [J]. IEEE Transactions on SMC-Part C: Applications and Review, 2001, 31 (4): 497-508.

[45] Cauwenberghs G, Poggio T. Advances in Neural Information Processing Systems 13 [M]. MIT Press, 2001.

[46] Chapelle O, Scholkopf B, Zien A. Semi-Supervised Learning [M]. The MIT Press, 2006.

[47] Cohen I, Sebe N, Cozman F G. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003: 595-604.

[48] Yao J, Zhang Z F. Semi-supervised learning based object detection in aerial imagery [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005: 1011-1016.

[49] Li L J, Wang G, Li F F. Optimol: automatic object picture collection via incremental model learning [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recogni-

tion, 2007: 604-611.

[50] Fergus R, Perona P, Zisserman A. Object class recognition by unsupervised scale-invariant learning [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003, 2: 264-271.

[51] Crandall D, Huttenlocher D. Weakly supervised learning of part-based spatial models for visual object recognition [C]. LNCS 3979: Proceedings of the European Conference on Computer Vision. Berlin / Heidelberg: Springer, 2006: 16-29.

[52] Duda R O, Hart P E, Stork D G. Pattern Classification: Second Edition [M]. New York: John Wiley & Sons, 2001.

[53] 肖海涛，张法瑞. 自然辩证法简编 [M]. 北京：北京航空航天大学出版社，1996.

[54] 边肇祺，张学工. 模式识别 [M]. 2 版. 北京：清华大学出版社，2000.

[55] 沈庭芝，方子文. 数字图像处理及模式识别 [M]. 北京：北京理工大学出版社，2000.

[56] 林开颜，吴军辉，徐立鸿. 彩色图像分割方法综述 [J]. 中国图象图形学报. 2005, 10 (1): 1-10.

[57] 刘陈. 知识驱动的图像对象分割方法及其应用研究 [D]. 北京：北京理工大学，2009.

[58] Wang J, Cohen M F. An iterative optimization approach for unified image segmentation and matting [C]. Proceedings of the IEEE International Conference on Computer Vision, 2005: 936-943.

[59] Boykov Y, Jolly M. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images [C]. Proceedings of the IEEE International Conference on Computer Vision, 2001: 105-112.

[60] Kumar M, Torr P, Zisserman A. OBJ CUT [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005: 18-25.

[61] Wang J, Ying Y, Guo Y, et al. Automatic foreground extraction of head shoulder images [C]. The 24th Computer Graphics International Conference, 2006: 385-396.

[62] Kohli P, Rihan J, Bray M, et al. Simultaneous Segmentation and Pose Estimation of Humans using Dynamic Graph Cuts [J]. International Journal of Computer Vision, 2008, 79 (3): 285-298.

[63] Borenstein E, Ullman S. Learning to segment [C]. Proceedings of the European Conference on Computer Vision, 2004: 315-328.

[64] Levin A, Weiss Y. Learning to combine bottom-up and top-down segmentation [C]. Proceedings of the European Conference on Computer Vision, 2006: 581-594.

[65] Shotton J, Winn J, Rother C, et al. TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context [J]. International Journal of Computer Vision, 2009, 81 (1): 2-23.

[66] Winn J, Shotton J. The layout consistent random field for recognizing and segmenting partially occluded objects [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006: 37-44.
[67] 章毓晋. 图像工程 [M]. 北京: 清华大学出版社, 2005.
[68] 高隽, 谢昭. 图像理解理论与方法 [M]. 北京: 科学出版社, 2009.
[69] 马莉, 范影乐. 纹理图像分析 [M]. 北京: 科学出版社, 2009.
[70] Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception [J]. IEEE Transactions on Systems, Man and Cybernetics, 1978, 8 (6): 460-473.
[71] Hu M K. Visual pattern recognition by moment invariant [J]. IRE Transactions on Information Theory, 1962, 1 (8): 179-187.
[72] Boggess A, Narcowich F J. 小波与傅里叶分析基础 [M]. 芮国胜, 等译. 北京: 电子工业出版社, 2004.
[73] Kohavi R, Tohu G. Wrappers for feature selection [J]. Artificial Intelligence, 1997, 97 (1-2): 273-324.
[74] Tibshirani R. Regression selection and shrinkage via the lasso [J]. Journal of the Royal Statistical Society, 1996, 58 (1): 267-288.
[75] Bi J, Bennett K, Embrecht M, et al. Dimensionality reduction via sparse support vector machines [J]. Journal of Machine Learning Research, 2003, 3: 1229-1243.
[76] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, 1: 511-518.
[77] Bins J, Draper B. Feature selection from huge feature sets [C]. Proceedings of the IEEE International Conference on Computer Vision, 2001, 2: 159-165.
[78] Jurie F, Triggs B. Creating efficient codebooks for visual recognition [C]. Proceedings of the IEEE International Conference on Computer Vision, 2005: 604-610.
[79] Agarwal S, Awan A, Roth D. Learning to detect objects in images via a sparse, part-based representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26 (11): 1475-1490.
[80] Mikolajczyk K, Leibe B, Schiele B. Local features for object class recognition [C]. Proceedings of the Tenth IEEE International Conference on Computer Vision, 2005, 2: 1792-1799.
[81] 托马斯. 信息论基础 [M]. 阮吉寿, 张华, 译. 北京: 机械工业出版社, 2008.
[82] Battiti R. Using mutual information for selecting features in supervised neural net learning [J]. IEEE Transactions on Neural Networks, 1994, 5 (4): 537-550.
[83] Fleuret F. Fast binary feature selection with conditional mutual information [J]. Journal of Machine Learning Research, 2004, 5: 1531-1555.

[84] Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy [J]. Journal of Machine Learning Research, 2004, 5: 1205-1224.

[85] Yang Y M, Pedersen J O. A comparative study on feature selection in text categorization [C]. Proceedings of the International Conference on Machine Learning, 1997: 412-420.

[86] Jolliffe I T. Principal Component Analysis: Second Edition [M]. New York: Springer-Verlag, 2002.

[87] Shlens J. A Tutorial on Principal Component Analysis [EB/OL]. (2003) [2004-10-27]. http://www. snl. salk. edu/ -shlens/pub/notes/pca. pdf.

[88] Hyvarinen A, Oja E. Independent component analysis: algorithms and applications [J]. Neural Networks, 2000, 13 (4-5): 411-430.

[89] Dai D Q, Yuen P C. Regularized discriminant analysis and its application to face recognition [J]. Pattern Recognition, 2003, 36 (3): 845-847.

[90] Yang J, Yang J Y. Why can LDA be performed in PCA transformed space [J]. Pattern Recognition, 2003, 36 (2): 563-566.

[91] Yu H, Yang J. A direct LDA algorithm for high-dimensional data——with application to face recognition [J]. Pattern Recognition, 2001, 34 (10): 2067-2070.

[92] Loog M, Duin R P W. Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26 (6): 732-739.

[93] Hild II K E, Erdogmus D, Tokkola K, Principe J C. Feature extraction using information-theoretic learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28 (9): 1385-1392.

[94] Zhu M, Matinez A M. Subclass discriminant analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28 (8): 1274-1286.

[95] He X F, Yan S C, Hu Y X, et al. Learning a locality preserving subspace for visual recognition [C]. Proceedings of the IEEE International Conference on Computer Vision, 2003: 385-392.

[96] Yan S C, Xu D, Zhang B Y, et al. Graph embedding and extension: a general framework for dimensionality reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29 (1): 40-51.

[97] Geng X, Zhan D C, Zhou Z H. Supervised nonlinear dimensionality reduction for visualization and classification [J]. IEEE Transactions on Systems, Man and Cybernetics, Part B, 2005, 35 (6): 1098-1107.

[98] Scholkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem [J]. Neural Computation, 1998, 10 (5): 1299-1319.

[99] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach [J]. Neural

Computation, 2000, 12 (10): 2385-2404.

[100] Lu J W Plataniotis K N, Venetsanopoulos A N. Face recognition using kernel direct discriminant analysis algorithms [J]. IEEE Transactions on Neural Networks, 2003, 14 (1): 117-126.

[101] Yang J, Zhang D, Frangi A F, Yang J Y. Two-dimensional PCA: a new approach to appearance-based face representation and recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26 (1): 131-137.

[102] Yang J, Zhang D, Yong X, Yang J Y. Two-dimensional discriminant transform for face recognition [J]. Pattern Recognition, 2005, 38 (7): 1125-1129.

[103] Kanal L N. Patterns in pattern recognition: 1968-1974 [J]. IEEE Transactions on Information Theory, 1974, 20 (6): 697-722.

[104] 阮秋琦. 数字图像处理学 [M]. 北京: 电子工业出版社, 2001.

[105] 贾云得. 机器视觉 [M]. 北京: 科学出版社, 2000.

[106] 程云鹏, 张凯院, 徐仲. 矩阵论 [M]. 西安: 西北工业大学出版社, 2006.

[107] Rubner Y, Puzicha J, Tomasi C, et al. Empirical evaluation of dissimilarity measures for color and texture [J]. Computer Vision and Image Understanding, 2001, 84 (1): 25-43.

[108] Aiyer A, Pyun K, Huang Y Z, et al. Lloyd clustering of gauss mixture models for image compression and classification [J]. Signal Processing: Image Communication, 2005, 20 (5): 459-485.

[109] Breiman L. Bagging predictors [J]. Machine Learning, 1996, 24 (2): 123-140.

[110] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of Computer and System Sciences, 1997, 55 (1): 119-139.

[111] Shapire R E. BoosTexter: A boosting-based system for text categorization [J]. Machine Learning, 2000, 39 (2-3): 135-168.

[112] Dietterich T G, Baliri G. Solving multiclass learning problems via error-correcting output codes [J]. Journal of Artificial Intelligence Research, 1995, 2: 263-286.

[113] 韩力群. 人工神经网络理论、设计及应用 [M]. 2 版. 北京: 化学工业出版社, 2007.

[114] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines [M]. Cambridge: Cambridge University Press, 2000.

[115] Ulusoy I, Bishop C M. Generative versus discriminative methods for object recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005, 2: 258-265.

[116] Holub A, Perona P. A discriminative framework for modeling object classes [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005: 664-671.

[117] Liu C L, Sako H, Fujisawa H. Discriminative learning quadratic discriminant function for

handwriting recognition [J]. IEEE Transactions on Neural Networks, 2004, 15 (2): 430-444.

[118] Lasserre J A, Bishop C M, Minka T P. Principled hybrids of generative and discriminative models [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006: 87-94.

[119] Grossman D, Domingos P. Learning Bayesian network classifiers by maximizing conditional likelihood [C]. Proceedings of the International Conference on Machine Learning, 2004.

[120] Greiner R, Su X Y, Shen B, et al. Structural extension to logistic regression: discriminative parameter learning of belief net classifiers [J]. Machine Learning, 2005, 59 (3): 297-322.

[121] Jain A K, Duin R P W, Mao J. Statistical pattern recognition: a review [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22 (1): 4-37.

[122] Dzeroski S, Raedt L D, Driessens K. Relational reinforcement learning [J]. Machine Learning, 2001, 43 (1-2): 7-52.

[123] Ormoneit D, Sen S. Kernel-based reinforcement learning [J]. Machine Learning, 2002, 49 (2-3): 161-178.

[124] Fawcett T. ROC graphs: Notes and practical considerations for data mining researchers [R]. Technical report HPL-2003-4. HP Laboratories, Palo Alto, CA, USA. 2003.

[125] Lachiche N, Flach P. Improving accuracy and cost of two-class and multi-class probabilistic classifiers using ROC curves [C]. Proceedings of the International Conference on Machine Learning, 2003: 416-423.

[126] Leibe B, Leonardis A, Schiele B. Combined object categorization and segmentation with an implicit shape model [C]. Proceedings of the European Conference on Computer Vision, 2004: 17-32.

[127] Leibe B, Leonardis A, Schiele B. Robust object detection with interleaved categorization and segmentation [J]. International Journal of Computer Vision, 2008, 77 (1-3): 259-289.

[128] 韩家炜, 坎伯. 数据挖掘: 概念与技术 [M]. 范明, 等译. 北京: 机械工业出版社, 2006.

[129] 宗成庆. 统计自然语言处理 [M]. 北京: 清华大学出版社, 2008.

[130] Vapnik V N. Statistical Learning Theory [M]. New York: John Wiley & Sons, 1998.

[131] Boughorbel S, Tarel J, Boujemaa N. The intermediate matching kernel for image local features [C]. Proceedings of International Joint Conference on Neural Networks, 2005, 2: 889-894.

[132] Revaud J, Lavoue G, Ariki Y, et al. Fast and cheap object recognition by linear combination of views [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007: 194-201.

[133] Ohba K, Ikeuchi K. Recognition of the Multi Specularity Objects Using the Eigen Window [C]. Proceedings of the International Conference on Pattern Recognition, Vienna, 1996: 692-696.

[134] Ohba K, Ikeuchi K. Detectability, Uniqueness, and Reliability of Eigen Windows for Stable Verification of Partially Occluded Objects [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19 (9): 1043-1047.

[135] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005: 886-893.

[136] Zhu Q, Avidan S, Yeh M, et al. Fast human detection using a cascade of histograms of oriented gradients [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006, 2: 1491-1498.

[137] Gool L, Moons T, Ungureanu D. Affine / photometric invariants for planar intensity patterns [C]. Proceedings of the European Conference on Computer Vision, 1996: 642-651.

[138] Belongie S, Malik J, Puzicha J. Shape context: A new descriptor for shape matching and object recognition [C]. Proceedings of the Neural Information Processing Systems, 2000: 831-837.

[139] Berg A, Berg T, Malik J. Shape matching and object recognition using low distortion correspondences [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005: 26-33.

[140] Fergus R, Perona P, Zisserman A. A sparse object category model for efficient learning and exhaustive recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005: 380-387.

[141] Beaudet P. Rotationally invariant image operators [C]. Proceedings of the International Joint Conference on Pattern Recognition, 1978: 579-583.

[142] Harris C, Steven M. A combined corner and edge detector [C]. Proceedings of the Conference on Alvey Vision Conference, 1988: 189-192.

[143] Mikolajczyk K, Schmid C. Indexing based on scale invariant interest points [C]. Proceedings of the Eighth International Conference on Computer Vision, 2001: 525-531.

[144] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors [J]. International Journal of Computer Vision, 2004, 60 (1): 63-86.

[145] Mikolajczyk K, Tuytelaars T, Schmid C, et al. A comparison of affine region detectors [J]. International Journal of Computer Vision, 2005, 65 (1-2): 43-72.

[146] Lowe D. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60 (2): 91-110.

[147] Kadir T, Zisserman A, Brady M. An affine invariant salient region detector [C]. Proceedings of the Europeon Conference on Computer Vision, 2004: 228-241.

[148] Matas J, Chum O, Urban M. Robust wide baseline stereo from maximally stable extremal regions [J]. Image and Vision Computing, 2004, 22 (10): 761-767.

[149] Nowak E, Triggs B. Sampling strategies for bag-of-features image classification [C]. Proceedings of the Europeon Conference on Computer Vision, 2006: 490-503.

[150] Koenderink J J. The structure of images [J]. Biological Cybernetics, 1984, 50: 363-396.

[151] Lindeberg T. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention [J]. International Journal of Computer Vision, 1993, 11 (3): 283-318.

[152] Lindeberg T. Scale-space theory: A basic tool for analyzing structures at different scales [J]. Journal of Applied Statistics, 1994, 21 (2): 224-270.

[153] Lowe D G. Object recognition from local scale-invariant features [C]. Proceedings of the IEEE International Conference on Computer Vision, 1998: 1150-1157.

[154] Mikolajczyk K, Schmid C. An affine invariant interest point detector [C]. Proceedings of the European Conference on Computer Vision, 2002: 128-142.

[155] Canny J. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8 (6): 679-698.

[156] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27 (10): 1615-1630.

[157] Lazebnik S, Schmid C, Ponce J. Sparse texture representation using affine-invariant neighborhoods [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003: 319-324.

[158] Ashbrook A, Thacker N, Rockett P, et al. Robust recognition of scaled shapes using pairwise geometric histograms [C]. Proceedings of the British Machine Vision Conference, 1995: 503-512.

[159] Ke Y, Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2004: 511-517.

[160] Papageorgiou C, Poggio T. A trainable system for object detection [J]. International Journal of Computer Vision, 2000, 38 (1): 15-33.

[161] Mohan A, Papageorgiou C, Poggio T. Example-based object detection in images by components [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23 (4): 349-361.

[162] Koenderink J, Doorn A, Representation of local geometry in the visual system [J]. Biological Cybernetics, 1987, 55: 367-375.

[163] Florack L, Romeny B, Koenderink J, et al. General intensity transformations and second order

invariants [C]. Proceedings of the Seventh Scandinavian Conference on Image Analysis, 1991: 338-345.

[164] Freeman W, Adelson E. The design and use of steerable filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13 (9): 891-906.

[165] Baumberg A. Reliable feature matching across widely separated views [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2000: 774-781.

[166] Schaffalitzky F, Zisserman A. Multi-view matching for unordered image sets [C]. Proceedings of the Europeon Conference on Computer Vision, 2002: 414-431.

[167] Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception [J]. IEEE Transactions on Systems, Man and Cybernetics, 1978, 8 (6): 460-473.

[168] 范立南, 韩晓微, 张广渊. 图像处理与模式识别 [M]. 北京: 科学出版社, 2007.

[169] Rosten E, Porter R, Drummond T. Faster and better: a machine learning approach to corner detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32 (1): 105-119.

[170] Lai S H, Fang M. A hybrid image alignment system for fast and precise pattern localization [J]. Real-Time Imaging, 2002, 8: 23-33.

[171] Smith S M, Brady J M. SUSAN - A New Approach to Low Level Image Processing [J]. International Journal of Computer Vision, 1997, 23 (1): 45-78.

[172] Trajkovic M, Hedley M. Fast corner detection [J]. Image and Vision Computing, 1998, 16 (2): 75-87.

[173] 王宇, 王涌天, 刘越. 基于 SIFT 和小波变换的图像拼接算法 [J]. 北京理工大学学报, 2009, 29 (5): 423-436.

[174] Brown M, Lowe D G. Automatic panoramic image stitching using invariant features [J]. International Journal of Computer Vision, 2007, 74 (1): 59-73.

[175] 郭耀, 张敏情, 杨晓元, 等. 基于小区域特征的图像检索方法 [J]. 计算机工程, 2009, 35 (7): 200-202.

[176] 韩笑. SIFT 特征在基于内容图像检索中的应用研究 [D]. 北京: 北京理工大学, 2009.

[177] Marr D, Poggio T. A computational theory of human stereo vision [J]. Proceedings of the Royal Society of London. Series B, 1979, 204 (1): 301-302.

[178] Friedman J H, Bentley J L, Finkel R A. An algorithm for finding best matches in logarithmic expected time [J]. ACM Transactions on Mathematical Software, 1977, 3 (3): 209-226.

[179] Su M S, Wang L, Cheng K Y. Analysis on multiresolution mosaic images [J]. IEEE Transactions on Image Processing, 2004, 13 (7): 952-959.

[180] 李晓明, 赵训坡, 郑链, 等. 基于 Fourier-Mellin 变换的图像配准方法及应用拓展 [J]. 计算机学报, 2006, 29 (3): 466-472.

[181] 魏雪丽, 张桦, 马艳洁, 等. 基于最大互信息的图像拼接优化算法 [J]. 光电子·激光, 2009, 20 (10): 1399-1402.

[182] 陈辉, 龙爱群, 彭玉华. 由未标定手持相机拍摄的图片构造全景图 [J]. 计算机学报, 2009, 32 (2): 328-335.

[183] 苏娟, 林行刚, 刘代志. 一种基于结构特征边缘的多传感器图像配准方法 [J]. 自动化学报, 2009, 35 (3): 251-257.

[184] Szeliski R. Video mosaics for virtual environments [J]. IEEE Computer Graphics and Applications, 1996, 16 (2): 22-30.

[185] Shum H, Szeliski R. Construction of panoramic mosaics with global and local alignment [J]. International Journal of Computer Vision, 2000, 36 (2): 101-130.

[186] The Getty. Art and Architecture Thesaurus Online [OL]. http://www.getty.edu/research/conducting_research/vocabularies/aat, 2008.

[187] 图像词典. 图像图片搜索引擎 [OL]. http://cn.gograph.com, 2008.

[188] 金大卫, 胡知元. 基于语义的图像检索应用研究 [J]. 武汉大学学报: 信息科学版, 2009, 34 (10): 1255-1259.

[189] 束鑫, 吴小俊, 潘磊. 一种新的基于形状轮廓点分布的图像检索 [J]. 光电子·激光, 2009, 20 (10): 1385-1389.

[190] 李勇. 基于内容的图像检索技术研究 [D]. 长春: 吉林大学, 2009.

[191] 沈燕然. 基于内容的图像检索和视频标注 [D]. 上海: 复旦大学, 2009.

[192] 杨红菊. 基于内容的图像检索方法研究 [D]. 北京: 北京理工大学, 2009.

[193] Dance C, Willamowski J, Fan L, et al. Visual categorization with bags of keypoints [C]. In ECCV international workshop on statistical learning in computer vision, 2004.

[194] Salton G, Wong A, Yang C S. A vector space model for automatic indexing [J]. Communications of the ACM, 1975, 18 (11): 613-620.

[195] Weber M, Welling M, Perona P. Towards automatic discovery of object categories [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2000, 2: 101-108.

[196] Opelt A, Fussenegger M, Pinz A, et al. Weak hypotheses and boosting for generic object detection and recognition [C]. Proceedings of the European Conference on Computer Vision, 2004: 71-84.

[197] Thureson J, Carlsson S. Appearance based qualitative image description for object class recognition [C]. Proceedings of the European Conference on Computer Vision, 2004: 518-529.

[198] Deselaers T, Keysers D, Ney H. Improving a discriminative approach to object recognition using image patches [C]. In DAGM of Pattern Recognition, 2005: 326-333.

[199] Zhang J, Marszalek M, Lazebnik S, et al. Local features and kernels for classification of tex-

ture and object categories: a comprehensive study [J]. International Journal of Computer Vision, 2007, 73 (2): 213-238.

[200] Biederman I. Recognition-by-components: A theory of human image understanding [J]. Psychological Review, 1987, 94: 115-147.

[201] Baerveldt A J. A Vision System for Object Verification and Localization Based on Local Features [C]. Proceedings of the European Workshop on Advanced Mobile Robots, 1999: 57-64.

[202] Dinesh R, Guru D S. Recognition of Partially Occluded Objects Using Perfect Hashing: An Efficient and Robust Approach [C]. Proceedings of the Canadian Conference on Computer and Robot Vision, 2005: 528-535.

[203] 周振环. 基于角点特征的形状识别 [J]. 计算机工程, 2007, 33 (6): 22-26.

[204] 王鹏伟, 吴秀清, 余珊. 基于角点特征和自适应核聚类算法的目标识别 [J]. 计算机工程, 2007, 33 (6): 179-184.

[205] Koenderink J J, Van Doom A J. The internal representation of solid shape with respect to vision [J]. Biological Cybernetics, 1979, 32 (4): 211-216.

[206] 王向军, 王研, 李智. 基于特征角点的目标跟踪和快速识别算法研究 [J]. 光学学报, 2007, 30 (2): 360-364.

[207] Hotelling H. Analysis of a complex of statistical variables into principal components [J]. Journal of Educational Psychology, 1933, 24: 417-441.

[208] Huttenlocher D P, Rucklidge W J. A multi-resolution technique for comparing images using the Hausdorff distance [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1993: 705-706.

[209] Shilane P, Min P, Kazhdan M and Funkhouser T. Princeton shape benchmark [S], 2004, Available: http://shape.cs.princeton.edu/benchmark/.

○ ISBN 978-7-111-38182-2
○ 策划编辑：吕　潇
○ 封面设计：赵颖喆

上架指导：计算机／图形图像

ISBN 978-7-111-38182-2

9 787111 381822 >

地址：北京市百万庄大街22号　邮政编码：100037
电话服务
社服务中心：(010)88361066
销售一部：(010)68326294
销售二部：(010)88379649
读者购书热线：(010)88379203
网络服务
门户网：http://www.cmpbook.com
教材网：http://www.cmpedu.com

封面无防伪标均为盗版

定价：39.80元